UNIVERSITY OF ALBERTA

Layers of Epigenetic Regulation in the Transposed Imprinted Genes in the Prader-Willi Syndrome Domain

by

Jason C. Y. Lau ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Medical Sciences - Medical Genetics

Edmonton, Alberta
Fall 2005

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-08670-X

Your file Votre référence
ISBN:
Our file Notre reference
ISBN:

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

Not all genes are inherited equal. For parts of the mammalian genome, the two alleles inherited from the parents are made non-equivalent by epigenetic means. One allele is transcribed preferentially over the other, which is based on the gender of the parent contributing that allele. Human chromosome 15q11-q13 and mouse chromosome 7C contain large, syntenic imprinted domains. Previous work has established a model by which some of these genes are imprinted in a transcription dependant manner initiating at the imprinting center (IC). The transposed imprinted genes, *NDN/Ndn*, *MAGEL2/Magel2*, *MKRN3/Mkrn3* and *Frat3* are also paternally expressed and under control of this IC, although the mechanisms of this control have been less characterized.

To study imprinting of the transposed imprinted genes, a strategy was employed to examine tissue and allele-specific regulation at different scales. *NDN/Ndn* serves as an excellent representative for this cluster of co-regulated genes. Comparative analysis of the promoter sequences across different species gave clues as to the motifs involved in regulation of *NDN/Ndn*. Detailed examination of human *NDN* indicated tissue and allele-specific differences in accessibility of the promoter to *trans*-acting factors. The basis of this difference may lie in the differential chromatin context as evidenced by differences in histone modifications. Using chromatin immunoprecipitation (ChIP), I found that certain modifications correlated with expression of *NDN*, and that other marks are associated with allelic identity. In the search for a mechanism that potentially leads to allelic differences in chromatin modifications that exist regardless of tissue type, the involvement of a protein already shown to be important in imprinting was ascertained. This protein, CTCF, binds at

positions flanking the transposed imprinted genes in the mouse, and more interestingly, does so in an allele-specific manner.

These studies provide important clues about the layers of regulatory mechanisms in the tissue and allele-specific regulation of *NDN/Ndn* and the other transposed imprinted genes. From this work, I propose a model of regulation for the transposed imprinted genes that involves CTCF associated chromatin changes that lead to differential histone modifications and ultimately, to accessibility and function of individual promoters to bring about correct spatiotemporal and imprinted expression.

ACKNOWLEDGEMENTS

Throughout this adventure into the mysteries of genetics that is my PhD, I have been indebted to many whom I would like to thank.

Dr. Rachel Wevrick has instilled in me the fundamentals of science and provided an environment in which I had the freedom to put it into practice. She has nurtured my growth every step of the way and this work would not be possible without her guidance. Members of the lab, past and present, and many of whom have become a second family; also deserve a great deal of kudos for their support over the years. I would also like to thank my co-supervisor Dr. Alan Underhill and everyone in the department of Medical Genetics at the University of Alberta.

Most importantly, I would like to thank my friends and family. My father, mother and brother have always supported my pursuits and their faith in me gave me the confidence to aim for any height. Finally, I would like to dedicate this thesis to the memory of my mother who gave me the will to achieve.

Financial Support

I have been supported by the Natural Science and Engineering Research Council, the Alberta Heritage Foundation for Medical Research and the Faculty of Graduate Studies. My work has been supported by the Canadian Institutes of Health Research to RW.

TABLE OF CONTENTS

CHAPTER 1 ♦ STATE OF THE ART **1**

GENOMIC IMPRINTING 2

H19/Igf2 .... *5*

BWS and the Kcnq1 locus *9*

PWS cluster at 15q11-13 .... *10*

Long IC transcript model 22

Transposed imprinted genes 26

EPIGENETICS AND CHROMATIN CONTROL 28

DNA methylation *28*

Histone modifications *31*

Epigenetic interaction *49*

Epigenetic mechanisms in genomic imprinting *52*

CTCF BIOLOGY 54

CTCF, the mammalian insulator protein *54*

From insulator to "imprintor" *56*

OBJECTIVES AND RATIONALE 58

CHAPTER 2 ♦ MATERIALS AND METHODS **60**

Sequence analysis *61*

Transcription factor binding prediction 61

CTCF binding cluster prediction 61

Tissues and cultured cell lines *62*

DNA extraction and PCR *63*

RNA extraction and RT-PCR *63*

Nuclear extracts *64*

Electromobility shift assay (EMSA) *64*

Sodium bisulfite sequencing *64*

Chromatin immunoprecipitation (ChIP) *65*

ChIP with human fibroblasts and lymphocytes 65

ChIP with mouse tissues 69

Modification of measurement technique 70

CHAPTER 3 ♦ ANALYSIS OF THE *NDN* PROMOTER 75

Introduction 76

Footprinted sequences in the NDN promoter *77*

Results 79

Phylogenetic analysis of the NDN/Ndn promoter *79*

Verification of bioinformatically predicted PAX5 site by EMSAs *86*

Allelic differences in footprints are due to chromatin states *89*

Discussion 94

CHAPTER 4 ♦ TISSUE-SPECIFIC AND IMPRINTED HISTONE MODIFICATIONS OF THE HUMAN *NDN* GENE 99

Introduction 100

Results 102

Histone modification of NDN. *102*

Histone modification of the IC. *109*

Discussion 110

CHAPTER 5 ♦ CTCF BINDS DIFFERENTIALLY METHYLATED REGIONS IN THE IMPRINTED MOUSE PRADER-WILLI SYNDROME LOCUS 116

INTRODUCTION 117

RESULTS 119

Identification of potential CTCF binding sites in the mouse PWS/AS region *119*

Verification of CTCF binding in vivo by chromatin immunoprecipitation *127*

Histone modification at putative CTCF cluster sites *130*

DNA methylation patterns of Frat3, Mkrn3 and iCT-A CTCF cluster sites *137*

DISCUSSION 141

CHAPTER 6 ♦ CONCLUSIONS AND OPEN QUESTIONS 151

Regulation of the transposed imprinted genes *153*

Early events of the transposed imprinted genes *159*

Talking to the transposed imprinted genes *163*

REFERENCES 168

LIST OF TABLES

Table 1-1. Brno nomenclature for histone modifications. 35

Table 1-2. Summary of modifications. 36

Table 2-1. Table of oligonucleotides. 73

Table 3-1. Table of predicted TFs. 84

LIST OF FIGURES

Figure 1-1. Mechanism of imprinting at *Igf2/H19*. 7

Figure 1-2. Genomic organization of the human PWS/AS imprinted domain. 12

Figure 1-3. Upstream exons of *SNURF-SNRPN* and patient microdeletions. 15

Figure 1-4. Second functional IC in the mouse PWS/AS region. 19

Figure 1-5. Interaction between PWS-SRO and AS-SRO. 19

Figure 1-6. Long paternal transcript through telomeric PWS/AS region. 25

Figure 1-7. Levels of chromatin compaction. 33

Figure 1-8. Covalent modifications of canonical histone subunit tails. 35

Figure 1-9. Models of histone conservation during replication. 39

Figure 1-10. Mechanism of transmission of H3.3 states through replication. 46

Figure 1-11. Examples of interactions between modifications. 48

Figure 1-12. Binary switch theory. 48

Figure 1-13. Interaction between DNA methylation and histone modification. 51

Figure 1-14. Variation in ZF domain usage in DNA recognition. 55

Figure 2-1. Chromatin immunoprecipitation (ChIP) 68

Figure 2-2. Change in measurement methods. 72

Figure 3-1. Sequence alignment of *NDN/Ndn*. 80

Figure 3-2. Percentage identity plot and predicted TF sites. 81

Figure 3-3. Predicted factor binding sites in the transcription-related footprint. 85

Figure 3-4. EMSA of footprinted site to related paired domains. 88

Figure 3-5. RT-PCR of *PAX* genes in fibroblast lines. 88

Figure 3-6. EMSAs with patient cell line nuclear extracts. 91

Figure 3-7. RT-PCR of candidate factors binding to the *NDN* promoter. 93

Figure 4-1. *NDN* and surrounding regions. 105

Figure 4-2. Examples of qualitative and quantitative ChIP analysis. 106

Figure 4-3. Histone modifications of active transcription. 107

Figure 4-4. Preliminary quantitative data on ChIP in PWS and AS fibroblast lines. 108

Figure 4-5. ChIP data with antibodies against various histone modifications at SNRPN.... 108

Figure 4-6. Model in which histone methylation plays an early role in imprint establishment and maintenance in NDN. 114

Figure 5-1. Predicted CTCF binding clusters. 122

Figure 5-2. CTCF ChIP in mouse telomeric PWS/AS region. 129

Figure 5-3. Histone modifications at predicted CTCF clusters. 133

Figure 5-4. Expression of transposed imprinted genes. 136

Figure 5-5. CpG sites analyzed for DNA methylation. 139

Figure 5-6. DNA methylation of *Frat3*, *Mkrn3* and iCT-A. 140

Figure 5-7. Model of CTCF function at transposed imprinted genes. 145

Figure 5-8. Models of imprinting operating at PWS/AS domain. 150

Figure 6-1. Model of epigenetic relationship between imprinted and tissue-specific regulation. 157

Figure 6-2. Cycle of imprinted inheritance of the IC. 161

Figure 6-3. Models of distant epigenetic influence. 164

LIST OF ABBREVIATIONS, SYMBOLS AND NOMENCLATURE

AS – Angelman Syndrome
BWS – Beckwith Wiedemann Syndrome
ChIP – chromatin immunoprecipitation
DMD – differentially methylated domain
DMR – differentially methylated region
EMSA – electrophoretic mobility shift assay
FB – fibroblast cell line
FISH – fluorescence *in situ* hybridization
IC – imprinting center
ICR – imprint control region
HAT – histone acetyltransferase
HDAC – histone deacetylase
HMT – histone methyltransferase
LCL – lymphoblast cell line
LCR – locus control region
MBD – methyl binding domain
PBS – phosphate buffered saline
PWS – Prader-Willi Syndrome
RNAi – RNA interference
SRO – smallest region of overlap
TF – transcription factor
UPD – uniparental disomy
ZF – zinc finger

Chapter 1 ♦ State of the Art

Genomic Imprinting

It is fortunate that Gregor Mendel chose the humble pea plant for his Treatises on Plant Hybrids of 1865. The observations with which he discovered his three laws of heredity form the basis of genetics as we know it today. However, had his focus been on a mammalian model, imprinted inheritance may have wreaked havoc on his theories.

Genomic imprinting is a phenomenon by which alleles of certain genes are regulated differentially depending on the parental gender of origin (da Rocha and Ferguson-Smith 2004). For most genes in a mammalian genome, parental contributions of each allele are not overtly identified with respect to the parental gender from which it originated. Both alleles are free to exert their influence according to Mendelian laws. Imprinted genes, on the other hand, are inherited genetically from both parents, but each carries an epigenetic memory of the gender of the previous generation. This allelic identity causes the silencing of one allele in the offspring and breaks Mendel's first law by causing a functional hemizygous state for imprinted loci (independent dominance and recessiveness of alleles are made irrelevant).

For example, a trait that depends on an imprinted locus being expressed from the paternal allele and silent on the maternal allele (maternally imprinted) can appear to "skip" generations indefinitely through the maternal side of a family and reappear only when passed through a male germline. Thereafter, it can manifest itself if passed through the paternal side. This inactivation and activation of imprinted genes is done without changes in DNA sequence. Instead, allelic identity is kept epigenetically, that is, by modifications carried on or in association with DNA itself. These epigenetic marks are erased and written differentially in oocyte and sperm precursors, giving rise to differential allelic identity in the offspring.

As far back as 1965, inheritance and effects of mutation in an imprinted gene, *Igf2r* were characterized (Johnson 1974). The gene responsible had not been identified and was called the *T* locus, and instead of imprinting, the observations were considered within the nebulous phenomenon given the label of maternal effects (Haig 2004). Subsequently, debate raged as to the mechanistic defect in the *T* locus mutation; was it a cytoplasmic defect of the egg or maternal pronuclear defect? This mirrored a similar debate prior to 1984 regarding why parthenogenotes (embryos with uniparental genetic contribution) fail in development; was it a cytoplasmic deficiency in contribution from the sperm, or a nuclear genetic defect of the genome wide homozygous state? Both debates were settled in back-to-back papers in the April 5th issue of Nature, showing non-equivalence of parental genomes (Surani et al. 1984). The nuclear defect of the *T* locus lethality was also attributed to differential modification of parental genomes, but more specifically, of the locus on Chromosome 17 (McGrath and Solter 1984b). The two groups delineated imprinting as a field of study with seminal work on parthenogenetic and androgenetic mouse embryos (McGrath and Solter 1984a; Surani et al. 1984). Surani et al. used activated haploid eggs and added pronuclei from fertilized eggs to produce gynogenotes, while McGrath and Solter transplanted pronuclei between fertilized eggs to produce gynogenotes and androgenotes. These invariably failed to develop to term and showed the requirement of paternal and maternal contributions for normal development. While “epigenetics” was not a term coined for its current use back then, these studies implied the most fundamental concepts of imprinted epigenetic memory of the previous generation, and the switch of that memory through the germline without genetic changes.

In humans, similar phenomena occur spontaneously in the form of hydatidiform moles and ovarian teratomas, containing only paternal and maternal genomes, respectively (Mowery-Rushton et al. 1996; da Rocha and Ferguson-Smith 2004). Hydatidiform moles consist of extraembryonic-like tissues while ovarian teratomas resemble tissues from the three germ layers. Histologically, these human uniparental tissues give telltale signs of a possible reason for the evolution of genomic imprinting, the sexual conflict model (Haig and Graham 1991). This model posits that the evolutionary driving force for fixation of imprinting is due to the unique maternal-offspring relationship in mammals. The disparity between maternal and paternal resources spent on a placental fetus pressures the paternally derived genome to increase growth of the fetus at the cost of future fecundity of the female, while the maternal genome has an interest in keeping growth in check for the sake of future fecundity. Consistent with this, hydatidiform moles consisting of paternal contributions are biased towards extraembryonic tissue growth, which would later increase nutrient transfer from mother to fetus.

Also of great interest is the mechanism by which imprinting arose. Similarities to X-inactivation are undeniable, and parallels have given great insight into the functions of both processes. X-inactivation is the process by which one X-chromosome in female cells is inactivated to achieve dosage compensation with respect to the male complement of a single X (Lyon 1961). While random X-inactivation and relevance to disease has received the most attention, it may be the exception and not the rule as X-inactivation is imprinted in marsupials as well as extraembryonic tissues of some mammals. It has been postulated that the ancestral form of X-inactivation is imprinted, and that autosomes gained this characteristic through translocations from the X, and gave rise to autosomal imprinting (Lee 2003). Perhaps ancestral forms of one of the following imprinted genes or clusters will be found on the X-chromosome of a mammalian/marsupial ancestor. Debates continue regarding the role of imprinting in evolution, but it is clear numerous loci in the mammalian genome exhibit this phenomenon (Morison and Reeve 1998).

H19/Igf2

No comprehensive discussion of imprinting can omit mention of the *H19/IGF2* imprinted genes. They have become the prototype of all imprinted genes and is the most thoroughly characterized (Arney 2003; Delaval and Feil 2004). Located on mouse chromosome 7 and human chromosome 11, *H19* and *IGF2/Igf2* are reciprocally imprinted, being expressed from the maternal and paternal alleles, respectively (Figure 1-1 A) (Bartolomei et al. 1991; DeChiara et al. 1991). Imprinting of these two genes is controlled by an imprint control region (ICR) located between the two genes. Deletion of this *cis*-acting element results in loss of imprinting and biallelic expression of both genes. Consistent with the conflict theory of imprinting discussed above, *Igf2* is a growth factor, and biallelic expression results in an increase in growth (Thorvaldsen et al. 1998), consistent with previous knock-out studies (DeChiara et al. 1991). The function of the ICR depends on a differentially methylated region (DMR, or sometimes DMD for differentially methylated domain), which is methylated on the paternal allele and unmethylated on the maternal allele. A methylation-regulated boundary model has been postulated whereby methylation of the DMR upstream of *H19* allows enhancers to activate *Igf2* on the paternal allele and hypomethylation of the DMR activates a boundary element precluding enhancers from activating *Igf2*, which are now accessible to *H19* (Figure 1-1 A) (Schmidt et al. 1999; Delaval and Feil 2004). There are other allelic differences thought to contribute to allelic expression. Histone modifications for example have been found to be differential, with the unmethylated maternal allele of the ICR enriched for active modifications such as acetylation (Hu et al. 1998) and histone methylation at lysine 4 of H3, while the paternal allele carried histone methylation at lysine 9 of H3, a modification indicative of silent chromatin (Yang et al. 2003). Nuclease sensitivity, a measure of chromatin conformation also shows a more accessible maternal allele at the DMR (Hark and Tilghman 1998). Replication timing has also been found to be differential between the two alleles (Bickmore and Carothers 1995).

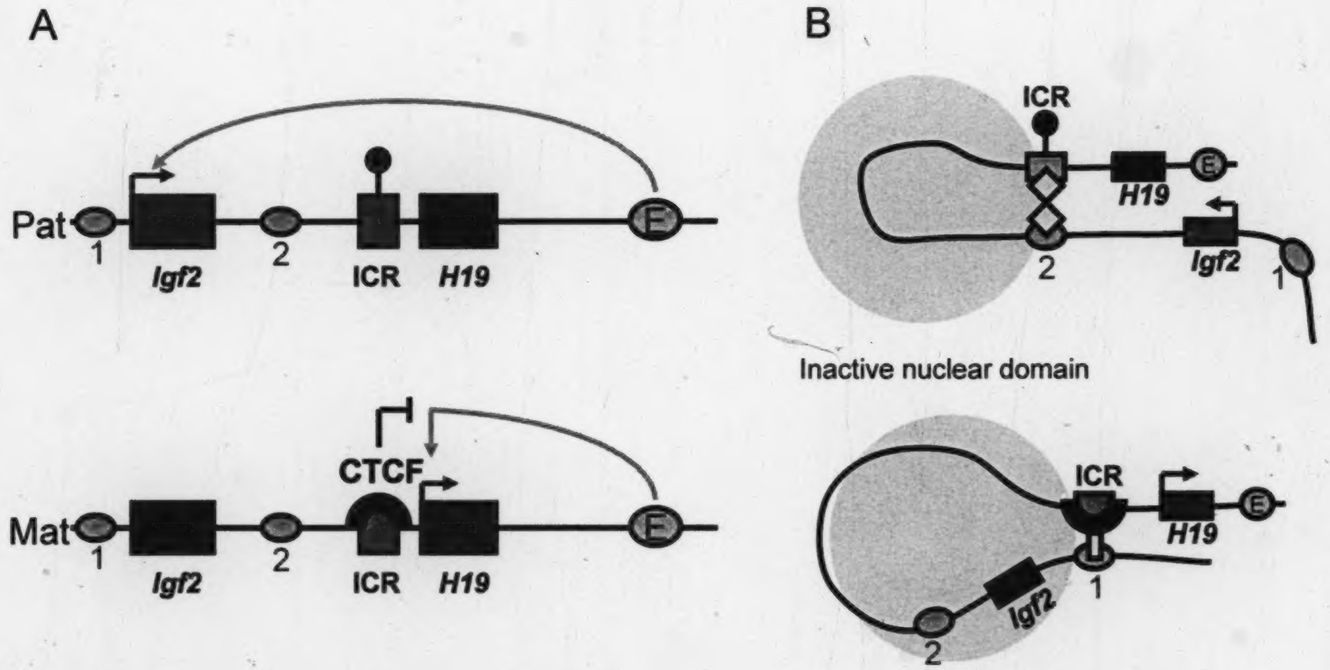

A
Pat
1
Igf2
2
ICR
H19
E
Mat
CTCF
1
Igf2
2
ICR
H19
E
B
ICR
H19
E
2
Igf2
1
Inactive nuclear domain
ICR
H19
E
1
Igf2
2

Figure 1-1. Mechanism of imprinting at *Igf2/H19*.
(A) Schematic of the methylation-regulated boundary model. *Igf2* out-competes *H19* for function of a distal enhancer on the paternal allele. CTCF is able to bind the unmethylated ICR of the maternal allele and block the enhancer from *Igf2*, allowing it to function at *H19*. (B) Intralocus loops in *Igf2/H19* imprinting. The methylated paternal ICR associates with DMR2 of *Igf2* to facilitate its expression. The CTCF bound maternal ICR associates with DMR1 of *Igf2* and shifts it to a silent nuclear domain.

Much has been learned since about the mechanisms of *H19/Igf2* imprinting, and this model has been elaborated in recent years. Most significantly, it has been found that CTCF, a protein that participates at almost all characterized mammalian insulators (Bell et al. 1999), functions in *H19/Igf2* imprinting (Bell and Felsenfeld 2000; Hark et al. 2000). Within the hypersensitive sites of the DMR (Hark and Tilghman 1998), are 4 CTCF binding sequences. These were found to have insulating activity in reporter assays, and binding was abrogated by DNA methylation. By binding the unmethylated maternal allele, CTCF acts to insulate *Igf2* from enhancers downstream of *H19*, thereby fulfilling the enhancer competition model mechanistically (Figure 1-1 A).

Furthermore, the mechanism of locus organization has also been elucidated. In elegant experiments involving two complementary lines of evidence, the higher order structure of the imprinted *H19/Igf2* genes have been characterized (Murrell et al. 2004). The data showing intralocus association imply that allele-specific long range loops are formed between the *H19* DMR and two other DMRs located at the 5' and 3' end of *Igf2* (DMR1 and DMR2, respectively). The model is that the *H19* DMR associates with DMR1 and DMR2 on the maternal and paternal alleles respectively, forming allele-specific epigenetic switch that shifts *Igf2* in and out of silent chromatin domain (Figure 1-1 B). How this aids in the access of enhancers in separated loop domains remains unclear and satisfactory convergence with the methylation-regulated boundary model is still lacking (Kato and Sasaki 2005).

BWS and the Kcnq1 locus

Adjacent to the *H19/Igf2* imprinted genes, lies another imprinted locus. Beckwith-Wiedemann Syndrome (BWS) results from paternal uniparental disomy (UPD) of chromosome 11p15, disrupting both imprinted regions (Weksberg et al. 2003). Genes in both regions are candidates for phenotypes in BWS; therefore, to avoid confusion, this second imprinted locus will herein be referred to as the *Kcnq1* locus. Five genes are imprinted in this locus: *TSSC3*, *SLC22A1L*, *CDKN1C*, *KCNQ1* and *LIT1* (seven in mouse with addition of *Mash2* and *Tssc4*). With the exception of *LIT1*, all are maternally expressed. The ICR for this cluster is the KvDMR, located in an intron of *KCNQ1*, and this ICR controls expression of *LIT1*, which is an antisense transcript to *KCNQ1* (Smilinich et al. 1999; Horike et al. 2000; Fitzpatrick et al. 2002). Imprinting of the *Kcnq1* locus is independent of the *H19/Igf2* locus (Caspary et al. 1998; Fitzpatrick et al. 2002). The mechanism of imprinting also appears to be quite different than this neighboring locus. Expression patterns do not support an enhancer competition model (Mancini-DiNardo et al. 2003). It has been shown however, that the unmethylated KvDMR functions as *LIT1*'s promoter and down-regulates other imprinted genes in *cis* on the paternal allele. The antisense nature of *LIT1* and the observation that *LIT1* is the only transcript expressed only on the paternal allele is reminiscent of the relationship between *XIST* and X-inactivation (Delaval and Feil 2004). In this system, *XIST* is transcribed from the silent X, and physically coats and silences that X in *cis* (Shibata and Lee 2004). Whether a similar mechanism exists for *LIT1* is unclear, however it is tempting to speculate that such a mechanism exits. Even within these neighboring examples of imprinted domains on the same chromosome, it is clear major differences exist in the mechanism of allelic regulation. Other imprinted domains such as *Igf2r, Callipyge* (not discussed here, but for review, see (Haig 2004)) and the PWS cluster, share this common theme of uncommon mechanisms.

PWS cluster at 15q11-13

A large cluster of imprinted genes resides on human chromosome 15q11-q13, with a region of conserved synteny on mouse chromosome 7C. In human, disruption of this region results in Prader-Willi syndrome (PWS) or Angelman syndrome (AS) (Nicholls and Knepper 2001). The determining factor between the two syndromes is parental origin of the remaining chromosome or region (Knoll et al. 1989). PWS results from a lack of paternal genetic contribution of the region either through maternal UPD (20%-30% of cases) or paternal deletion (~4Mb, 65%-75% of cases). In up to 5% of cases, an imprinting mutation causes assignment of a maternal identity to the paternal allele in the male germline. Angelman results from lack of maternal contribution by reciprocal chromosomal mechanisms as PWS, or by mutations in *UBE3A* (10%), a maternally expressed gene responsible for most if not all the features of AS (Burger et al. 2002). No single gene responsible for the neurobehavioral features of PWS, such as hypotonia, developmental delay and hyperphagia (Holm et al. 1993), has been identified (Lee and Wevrick 2000). The ICR for this region, and where imprinting mutations occur, lies in a region at the 5' end of *SNURF-SNRPN*, a bicistronic transcript encoding a small protein (SNURF) and a subunit (SMN) of a ribonucleoprotein (Sutcliffe et al. 1994; Gray et al. 1999). Deletion of this ICR and associated CpG island causes loss of imprinting for the entire region, including *SNURF-SNRPN*, *IPW* (Wevrick et al. 1994), a collection of snoRNAs (Cavaille et al. 2000; de los Santos et al. 2000), *NDN* (MacDonald and Wevrick 1997), *MAGEL2* (Lee et al. 2000), *MKRN3* (Jong et al. 1999), plus *Frat3* in mouse (Chai et al. 2001; Kobayashi et al. 2002) which are paternally expressed, and *UBE3A* (Rougeulle et al. 1997) and *ATP10C* (Meguro et al. 2001), which are maternally expressed (Figure 1-2).

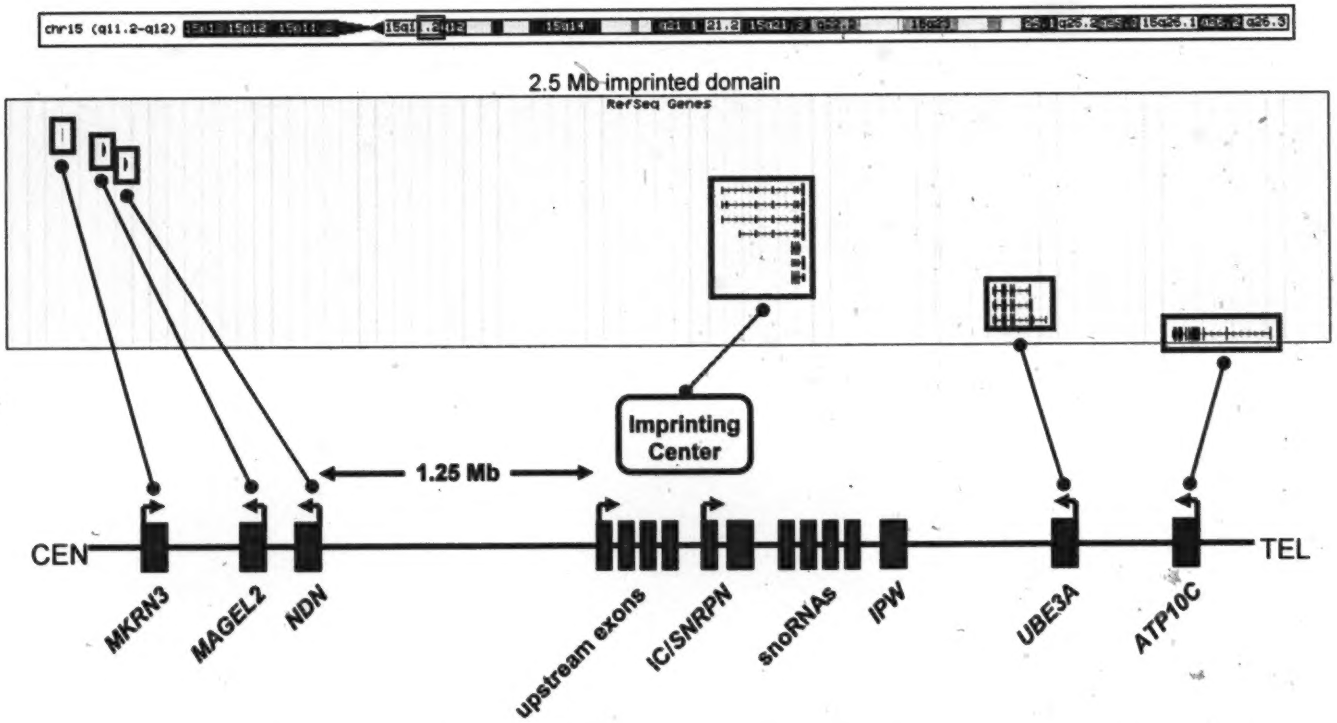

chr15 (q11.2-q12)
15p12
15q14
15q26.1
q26.3
2.5 Mb imprinted domain
RefSeq Genes
Imprinting Center
1.25 Mb
CEN
TEL
MKRN3
MAGEL2
NDN
upstream exons
IC/SNRPN
snoRNAs
IPW
UBE3A
ATP10C

Figure 1-2. Genomic organization of the human PWS/AS imprinted domain.
Chromosome 15 is shown at top, with the imprinted domain indicated by the red box. The 2.5 Mb domain is shown to scale in the center frame with transcripts indicated in blue. Note the large distance between the IC and the transposed imprinted genes. Genes are indicated at the bottom as a schematic. The IC is coincident with *SNRPN* exon 1.

The function of this ICR, more commonly referred to as the imprinting center (IC) at the PWS/AS locus, has been carefully dissected using a variety of patient studies, transgenic mice and other experiments. It was clear from UPD patients that PWS was due to a lack of paternally expressed genes instead of overexpression of maternally active genes from 15q11-13, since PWS resulting from maternal UPD had the same phenotype as paternal deletion (Glenn et al. 1996). A small number of cases are due to a heterogeneous class of imprinting mutations, either carrying microdeletions, which are often familial, and what are thought to be developmental or stochastic failures in the imprint process, which are sporadic (Nicholls et al. 1998). These microdeletions define an IC that lies at the 5' end of the *SNRPN* gene, where upstream exons not part of the protein coding potential of *SNRPN* have been detected (Figure 1-3) (Sutcliffe et al. 1994). These are referred to as the IC transcripts and they are important to the imprinting process as a splice mutation has been found in an AS patient (Dittrich et al. 1996). The microdeletion patients also delineate a bipartite functional structure for the PWS/AS IC, where a collection of AS patients define a smallest region of overlap (SRO), necessary for normal paternal to maternal allele identity switch, and a set of PWS patients define the PWS-SRO, necessary in the maternal to paternal switch (Figure 1-3) (Buiting et al. 1995; Saitoh et al. 1996). For example, a male inheriting a microdeletion of the PWS-SRO from his mother will not be able to reassign that allele to a paternal epigenotype for his progeny, who will have a 50% chance of inheriting this grandmaternal epigenotype from this male and develop PWS. A female on the other hand, inheriting the same microdeletion from her father will be able to reassign the epigenotype to a maternal one. Male to male and female to female transmission does not require reassignment.

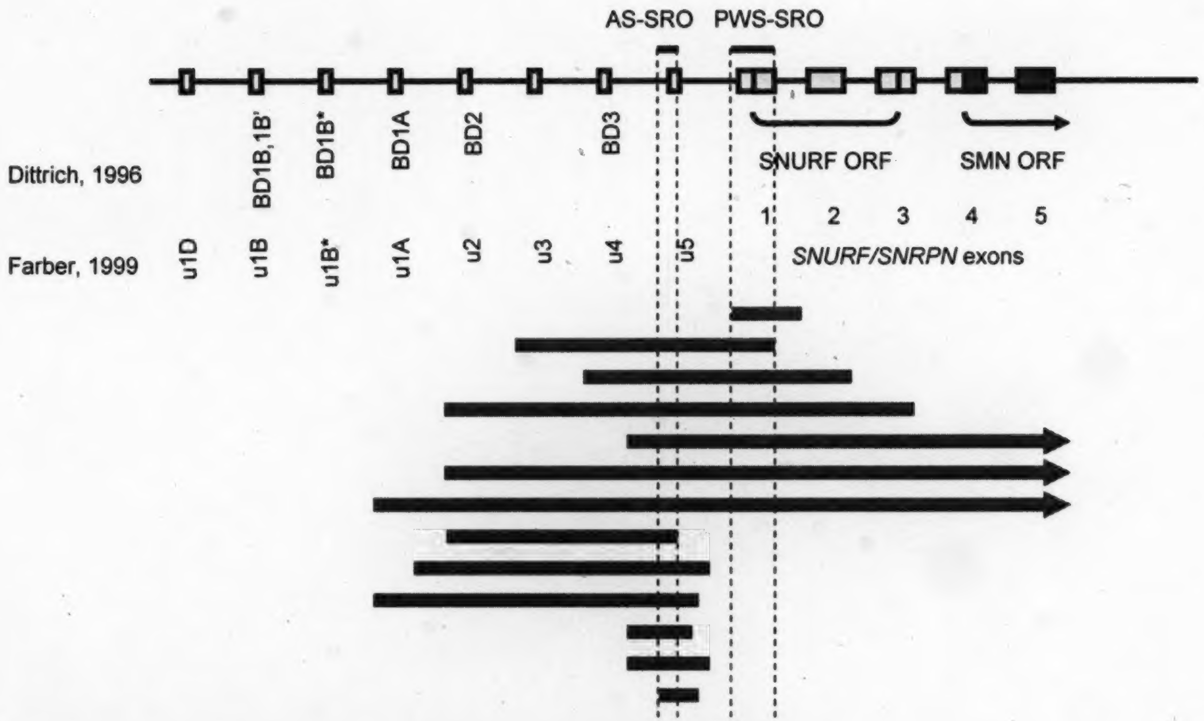

AS-SRO
PWS-SRO
Dittrich, 1996
BD1B,1B'
BD1B*
BD1A
BD2
BD3
SNURF ORF
SMN ORF
1
2
3
4
5
Farber, 1999
u1D
u1B
u1B*
u1A
u2
u3
u4
u5
SNURF/SNRPN exons

Figure 1-3. Upstream exons of *SNURF-SNRPN* and patient microdeletions.
At top is a schematic map (not to scale) of the upstream region of *SNURF-SNRPN*, with upstream exons. Nomenclature of these exons is provided below, with two previously published naming schemes. Bars below represent the extent of the microdeletions in patients that define the PWS (blue) and AS (red) SROs.

The region of conserved synteny in mouse chromosome 7C is likewise imprinted and behaves in much the same way. While useful in showing the conserved imprinted characteristics, the mouse models for the pathological phenotype were less clear. The PWS and AS models, achieved through maternal and paternal duplication of mouse 7C, did not recapitulate some aspects of the diseases (Cattanach et al. 1992; Cattanach et al. 1997). These were similar to a mouse strain carrying a fortuitous large deletion of the entire 7C imprinted domain (Gabriel et al. 1999). In retrospect, the difference in phenotype was not surprising, considering the number of genes involved and the potential that each may have subtle species-specific differences in function. However, the mouse system has been shown to be an excellent model for the study of imprinting in this region. Targeted deletions of part of *Snrpn* in a functional domain of the SmN protein showed that disruption of this gene had no effect on viability or imprinting, but a larger deletion involving the 5' end of *Snrpn* did have a phenotype (Yang et al. 1998). These mice shared the same failure to thrive phenotype as the maternal UPD mice (Cattanach et al. 1992). This phenotype is postulated to be causally related to the hypotonia seen in PWS neonates (Yang et al. 1998). More importantly, imprinting for the entire region was perturbed including dysregulation of imprinted genes more than a Megabase away.

How the PWS-SRO and the AS-SRO elements bring about the imprint switch for the locus was shown by recent work on the mechanisms by which these elements interact. While these two elements are 35kb apart endogenously, when brought together in close proximity, a transgenic construct is able to carry out all the steps of the imprinting process (Shemer et al. 2000). Using a series of transgenic mice with varying parts of the mouse equivalent of the PWS-SRO and the human AS-SRO, it was found that one kilobase of the human AS-SRO and 200bp of the mouse *Snrpn* minimal promoter was sufficient for imprinting of a reporter gene. In addition to paternal specific expression, the transgene also showed differential DNA methylation and asynchronous replication, recapitulating these features of the endogenous locus. Further dissection of the step-wise function of this IC construct has shown that the mouse PWS-SRO contains elements necessary for *de novo* methylation of the maternal allele in imprint establishment and elements necessary for maintenance of the imprinted state in somatic tissues during development (Kantor et al. 2004a). In addition, there is an element that prevents methylation on the paternal allele. Therefore, the PWS-SRO is not simply a locus control region being controlled by the AS-SRO, but carries signals necessary and specific to imprinting. While the AS-SRO is upstream of the PWS-SRO in certain aspects of imprinting, it is not a simple matter that the AS-SRO imprints the PWS-SRO, since the AS-SRO is not able to imprint an unrelated β-globin locus element (Shemer et al. 2000).

The situation in the mouse is complicated by the observation that a small 0.9 kb deletion of the region homologous to the human PWS-SRO does not perturb imprinting (Bressler et al. 2001), whereas a larger 35 kb deletion does (Yang et al. 1998; Chamberlain and Brannan 2001), suggesting *Snrpn* exon1 and the associated CpG island are not important in imprinting in the mouse. The explanation was revealed when it was discovered that the mouse has a second redundant imprinting center capable of functioning in the absence of the first (Figure 1-4) (Kantor et al. 2004b). A 4.8 kb deletion that deletes the first IC and part of the second exhibits partial imprinting defects (Bressler et al. 2001).

How the remnants of the second IC are able to carry out partial imprinting is not well characterized.

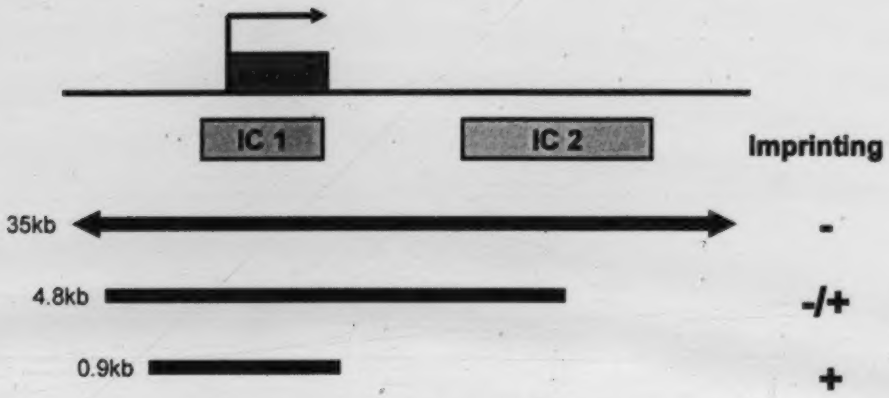

IC 1
IC 2
Imprinting
35kb
-
4.8kb
-/+
0.9kb
+

Figure 1-4. Second functional IC in the mouse PWS/AS region.
Schematic of the two redundant ICs equivalent to the PWS-SRO in the mouse. IC1 is coincident with exon 1 of *Snurf/Snrpn*, while IC2 is in the first intron. Below are mouse deletions of the IC, and their effects on imprinting (+, imprinting retained, -, imprinting disrupted) .

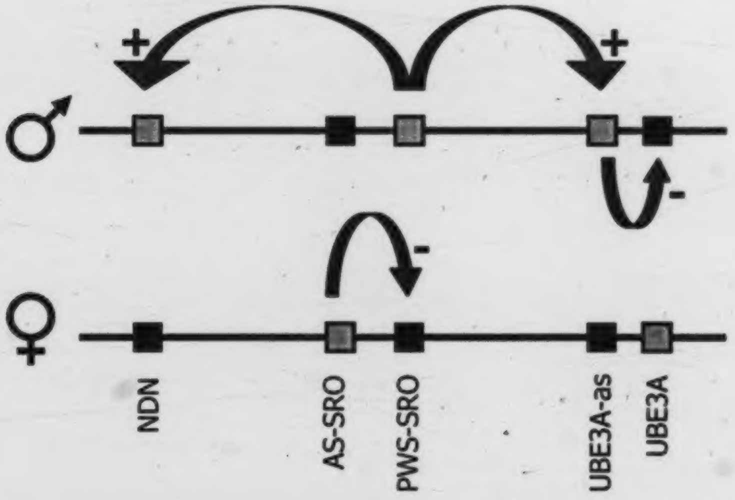

+
+
♂
-
-
♀
NDN
AS-SRO
PWS-SRO
UBE3A-as
UBE3A

Figure 1-5. Interaction between PWS-SRO and AS-SRO.
The PWS-SRO is believed to be a positive acting element on the imprinted domain, while the AS-SRO is believed to be a negative regulator of the PWS-SRO.

Further understanding of the *cis*-acting genetic epistasis between the two SRO elements was gained from studies using patient cells carrying microdeletions of the IC (Perk et al. 2002). It was found that the AS-SRO is associated with maternal allele-specific histone acetylation, H3 lysine 4 dimethylation (H3K4me2), and DNase I hypersensitivity, but surprisingly, not differential DNA methylation. In transgenic experiments however, differential DNA methylation are also observed (Kantor et al. 2004a). These epigenetic characteristics are not affected by deletion of the PWS-SRO in patients. This is in contrast to the reciprocal situation, whereby deletion of the AS-SRO causes biallelic hypomethylation, DNase I hypersensitivity and H3K4me2 modification at the PWS-SRO, features usually found only on the paternal allele, yet does not affect paternal allele-specific histone acetylation and asynchronous replication. From these observations, it is clear that the AS-SRO is necessary for the maternal methylation and certain other characteristics of the PWS-SRO, but does not depend on the PWS-SRO for its own epigenetic state. The PWS-SRO seems to be responsible for activation of paternal genes in *cis*, while the AS-SRO represses the PWS-SRO in *cis* on the maternal allele (Brannan and Bartolomei 1999) (Figure 1-5). This epistasis explains why deletion of both elements results only in PWS and not AS. When both are deleted, the paternal genes are not activated by the PWS-SRO on the paternal allele, while on the maternal allele, the AS-SRO is no longer required to repress the PWS-SRO.

As described, many genetic lesions lead to PWS and AS, but it is the epigenetic characteristics of this locus that are of interest here. The importance of epigenetic marking is demonstrated by a class of PWS and AS patients where no genetic lesions are found.

They are thought to carry epimutations, or errors in the epigenetic marking of alleles (Buiting et al. 2003). It is therefore critical to the understanding of PWS and AS imprinting to characterize the epigenetic regulation of the locus. In addition to DNA methylation, numerous lines of evidence indicate allele-specific differences in chromatin structure. Replication timing is often correlated to chromatin states in that euchromatin replicates earlier in S-phase than does heterochromatin (Goren and Cedar 2003). It has previously been observed that the two alleles of the PWS region replicate asynchronously (Knoll et al. 1994). Paternal-early and maternal-late replication was observed near the imprinted domain, while other patterns were observed more distally. DNase I hypersensitivity is commonly correlated with open chromatin (Weintraub and Groudine 1976). Two strong paternal allele-specific DNase I hypersensitivity sites flank exon 1 of *SNRPN*, while remaining resistant on the maternal allele (Schweizer et al. 1999). Interestingly, a less striking site of maternal allele-specific DNase I sensitivity was also observed coincident with the AS-SRO. Chromatin compaction as measured by density fractionation gave complementary results, where the paternal allele was measured to be less compact than the maternal allele (Watanabe et al. 2000). Larger scale organization has also been examined using two different fluorescence *in situ* hybridization (FISH) techniques, both indicating allele-specific patterns of matrix association (Greally et al. 1999; Kagotani et al. 2002).

The IC obviously plays an essential role at the PWS/AS imprinted domain. It is responsible for the correct expression of genes spread out over two Megabases, allelic identity and its maintenance throughout development. Certain details of how this genetic element receives the initial epigenetic mark have been elucidated. How this translates into long range regulation of genes in *cis* however is less clear.

Long IC transcript model

The genomic organization of the imprinted region between the IC and the maternally expressed gene *UBE3A* includes more than the aforementioned paternally expressed *IPW* and snoRNA genes and the maternally expressed *ATP10C*. There are also other transcripts such as the *PAR* (for Prader-Willi/Angelman region) transcripts of unknown function (*PAR-1, PAR-2, PAR-4, PAR-5, PAR-7* (Sutcliffe et al. 1994), *PAR-SN* (Ning et al. 1996)), some of which have been shown to be paternally expressed, and only in certain tissues such as brain and skeletal muscle, in contrast to the ubiquitous expression of *SNURF/SNRPN*. In addition, there is also a paternally expressed transcript in antisense orientation to *UBE3A* (Rougeulle et al. 1998) that is also restricted in tissue-specific expression patterns and is under imprinted control of the IC (*UBE3A-AS*, (Chamberlain and Brannan 2001)).

Study of the IC has always been daunting because of its transcriptional complexity. The basic structure of the *SNURF/SNRPN* gene includes ten exons and is unusual in that it encodes a bicistronic transcript with two open reading frames, one for SMN (exons 4-10) and the SNRPN upstream reading frame protein (*SNURF*, exons 1-3) (Dittrich et al. 1996; Gray et al. 1999). However, operons are rare in mammalian genomes and the significance of this bicistronic gene to imprinting is unclear. While exon 1 is associated with a differentially methylated CpG island and is the most commonly transcribed 5' end to the *SNURF/SNRPN* gene, many 5' upstream exons exist and are thought to function in imprinting (Dittrich et al. 1996; Farber et al. 1999). There are at least seven upstream exons in human, with several others that may be pseudo-exons (Figure 1-3). Most of these special U transcripts initiate at either u1B or u1A, exclude exon 1, splice directly to exon 2, and include variable numbers of the rest of the *SNRPN* exons (Farber et al. 1999). Most significantly, exon u5 is deleted in all AS patients with a microdeletion of the IC, suggesting this exon and perhaps upstream transcription plays a critical role in the paternal to maternal imprint switch. Also, whereas expression of *SNURF/SNRPN* is ubiquitous, transcripts containing these upstream exons are mainly in tissues such as brain, heart, testis and ovary (Dittrich et al. 1996). The significance of this will not become apparent until put into context of the rest of the transcription unit.

This transcription unit becomes more complex with addition of extra 3' exons into the picture. As with the 5' exons, each report provided evidence of more and more 3' exons. First were exons immediately downstream of *SNRPN* (10a, 11, 12) that connected to the *PAR* transcripts, previously thought to be individually expressed (Buiting et al. 1997), then to a larger set (13-20) that encompassed some of the snoRNAs (Wirth et al. 2001). It was also shown that some of these exons are only found in certain tissues that may not include *SNURF/SNRPN* exons (Buiting et al. 1997).

This transcriptional nightmare was brought into a single model with the realization that all of the paternally expressed transcripts between the IC upstream exons and *UBE3A-AS* represent alternative transcripts of a single transcription unit at least 460kb in size (Runte et al. 2001). This was first suggested by the fact that all the paternally expressed transcripts are transcribed from centromere to telomere whereas both maternal genes are transcribed in the opposite direction. Through the sequencing of multiple RT-PCR products and ESTs, a total of 128 new 3'exons were found distributed throughout the telomeric side of the human imprinted domain, linking pieces of previously known paternal transcripts. There was a high degree of alternative splicing for the products described, but many seemed to overlap the snoRNA genes, probably acting as host transcripts for these promoter-less intronic functional RNAs. The current model is that the paternal long transcript imparts paternal expression of genes in *cis* and in the same transcriptional direction, and downregulates the paternal allele of the maternally expressed gene *UBE3A* by an antisense mechanism since it is transcribed in the opposing direction (Figure 1-6). Another twist to the model was presented with evidence of the same long transcriptional unit in the mouse (Landers et al. 2004). In the mouse, it was observed that many of the alternative transcripts start with the upstream exons, some of which are 500kb upstream of *Snrpn* exon 1, making the transcription unit 1 Mb in length. Some of these transcripts also spliced in such a way as to exclude *Snrpn*. The most interesting lesson learned from the mouse is that these transcripts starting at the upstream exons and ending with *Ube3a-as* are tissue-specific, which provides an explanation as to how the imprinting of *Ube3a* is tissue-specific while *Snurf/Snrpn* is not. It is probable that this mechanism also operates in the human but this awaits experimental verification. How the maternal expression of *ATP10C* fits into this model is also not known, but it is possible further work will find exons extending telomeric to *UBE3A*.

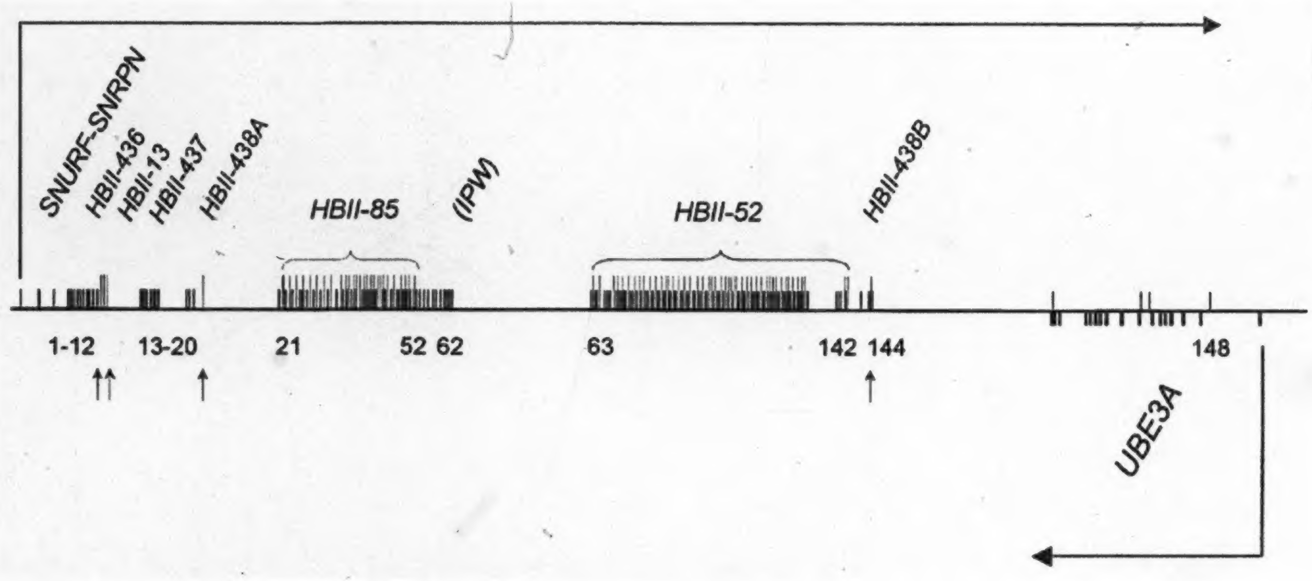

SNURF-SNRPN
HBII-436
HBII-13
HBII-437
HBII-438A
HBII-85
(IPW)
HBII-52
HBII-438B
1-12
13-20
21
52 62
63
142 144
148
UBE3A

Figure 1-6. Long paternal transcript through telomeric PWS/AS region.
The 148 identified exons part of the paternal transcript controlling imprinting of paternal and maternal genes telomeric to the IC. *HBII* genes are the human snoRNA genes. Reproduced from figure 5A, (Runte et al. 2001).

Transposed imprinted genes

While the long transcript model is a satisfying explanation for imprinting of the genes between the IC and *ATP10C*, and the tissue specificity of imprinting, it is not consistent with the imprinting of the paternally expressed genes *NDN* (MacDonald and Wevrick 1997), *MAGEL2* (Lee et al. 2000), and *MKRN3* (Jong et al. 1999) on the centromeric side of the IC (in human, telomeric side in mouse). The first inconsistency is the transcriptional direction of these genes, with *MKRN3* transcribed from centromere to telomere and *NDN* and *MAGEL2* transcribed in the opposite direction (Figure 1-2). In the mouse, an additional gene exists called *Frat3*, which is the product of a more recent transposition event (Chai et al. 2001; Kobayashi et al. 2002). All four genes seem to be processed forms of other genes, indicating they are likely products of past retrotransposition events. Therefore, they will be referred to as the transposed imprinted genes. All of these genes are associated with a differentially methylated region in their CpG islands, in contrast to the genes thought to be under control of the long antisense transcript. This fits well with the model that most maternally silenced genes are mediated by DNA methylation, and most paternally silenced genes are associated with an antisense mechanism (Reik and Walter 2001). A theory has been proposed to explain this difference that is based on the early demethylation events of the paternal genome in the zygote. Whereas the maternal genome demethylates passively, the paternal genome may have evolved other mechanisms such as antisense transcription to silence imprinted genes in spite of active demethylation (Mayer et al. 2000). Therefore, while the IC can silence the paternal alleles of the maternally expressed genes *UBE3A* and *ATP10C* via the long antisense transcript, a separate mechanism may exist for regulation of the paternally expressed genes such as the transposed imprinted genes, which carry individual differentially methylated CpG islands. In light of the long transcript, it is clear how the action of the IC can span the 460kb from the IC to *UBE3A* in its influence. Outside of this model, there is no proven way for the IC to influence the transposed imprinted genes at a distance. Characterization of the mechanisms of imprinting for these genes will be critical in the understanding of the function of the IC. This will be explored in detail in the following chapters. To understand the regulation of the transposed imprinted genes, it will be useful to study one as a model for the others in the cluster. Chapter 3 will focus on the immediate regulation of the promoter of *NDN*. Chapter 4 builds on the observations at the promoter by examining the chromatin context of the region containing *NDN*. Chapter 5 investigates the mechanism of the entire transposed imprinted domain and will reveal a possible model to reconcile the epigenetic regulation of imprinting and tissue-specific expression.

Epigenetics and Chromatin Control

The differential allelic regulation of imprinted genes is transmitted between generations and cell divisions without changes in DNA sequence. Memory of the parental origin and allelic identity is instead recorded in an epigenetic manner. DNA methylation, histone modifications and their interaction with each other are all involved in the orchestra of events that leads to correct imprinted gene regulation.

DNA methylation

Without a doubt, DNA methylation is the best studied epigenetic mark. It is truly an archetype of an epigenetic mark in that it sits atop DNA to carry extra information and it has a clear mechanism of heritability through its maintenance during DNA replication. It was not until the late 1970's when this minor base variant was associated with gene activity (Razin and Riggs 1980). Since then, its importance has been demonstrated in multiple systems such as tissue-specific regulation, differentiation, cancer, X-inactivation, and genomic imprinting (Paulsen and Ferguson-Smith 2001). Methylation of CpG dinucleotides of mammalian genomes is accomplished by DNA methyltransferases (DNMTs) that act on hemi-methylated DNA for maintenance (DNMT1) and Dnmts that act on unmethylated DNA to generate *de novo* methylation patterns (DNMT3A/B) (Bestor 2000). Mutations in *DNMT3B* are found in ICF syndrome, which shows immunodeficiency and centromeric instability on certain chromosomes (OMIM: 242860, (Hansen et al. 1999; Xu et al. 1999)). Targeted deletions of the *Dnmts* are lethal, with *Dnmt1*-nulls being the most severe, and affecting imprinted genes and X-inactivation (Li et al. 1992; Li et al.

1993; Beard et al. 1995; Okano et al. 1999). Over-expression of *Dnmt1* causes hypermethylation and loss of imprinting (Biniszkiewicz et al. 2002). Other DNMTs such as DNMT2 and DNMT3L exist, though they are without *in vitro* methylating activities, and may participate in pathways related to their true methyltransferase relatives (Bestor 2000).

DNA methylation is thought to modify protein-DNA interactions to bring about its biological effects (Razin and Riggs 1980). Many transcription factors are not able to bind if their target DNA is methylated (Tate and Bird 1993). The DNA binding factor CTCF has received much attention of late and has been shown to be sensitive to DNA methylation (Bell and Felsenfeld 2000; Hark et al. 2000). This will be discussed in a following section (page 54). DNA methylation does not always preclude binding of factors, and in fact, there are protein domains with specific affinity for methylated DNA (Hendrich and Tweedie 2003). There are five proteins with a methyl CpG binding domain (MBD); MeCP2, MBD1, MBD2, MBD3 and MBD4. Most have been shown to have a repressive activity, while MBD4 seems to function in a repair pathway that deals with the mutagenic cost of having methylated cytosine residues in the genome (Millar et al. 2002). MBD2 may also have direct DNA demethylase activity but this has been controversial (Bhattacharya et al. 1999; Ng et al. 1999). These proteins bring about transcriptional repression through recruitment of other proteins such as chromatin remodeling enzymes, histone deacetylases (HDACs) and histone methyltransferases (HMTs). MBD2 and MBD3 for example, are associated with the MeCP1 and NuRD repressive complexes respectively (Ng et al. 1999; Zhang et al. 1999). MeCP2 in particular has been associated with HDAC activities (Nan et al. 1998) as well as HMT activities (Fuks et al. 2003), and even associates with DNMT1 (Kimura and

Shiota 2003). Beyond that, much attention has been paid to *MECP2*, because it is a causative gene in Rett Syndrome (Bienvenu et al. 2000). Although widely expressed and thought to have a general role in gene repression, phenotypic defects are mainly neurological, perhaps due to sensitivity of the brain to loss of MeCP2 and perturbation in a subset of genes it regulates (Chen et al. 2001; Guy et al. 2001; Tudor et al. 2002). Most importantly to our discussion, there are isolated examples of loss of imprinting (LOI) in Rett Syndrome (Horike et al. 2005; Makedonski et al. 2005), although there are no obvious global defects in imprinting (Balmer et al. 2002).

Through the action of the maintenance methylase on newly synthesized DNA, methylation patterns can be transmitted through cell divisions. But in terms of imprinting, epigenetic signals must be transmitted through generations to record the parental gender of origin of an allele. This presents a challenge for DNA methylation as there is a global demethylation event early in zygotic development (Jaenisch 1997; Morgan et al. 2005). This challenge is especially daunting for the paternal genome since it is actively demethylated in the zygote after fertilization and before the first round of DNA replication. Fortunately, DMRs associated with ICs are able to retain their methylation state through an unknown mechanism and represent *bona fide* imprints from the last generation. However, many DMRs not part of ICs lose their methylation and must be re-established after the demethylation event (Hanel and Wevrick 2001). To combat the active demethylation of the paternal genome by the oocyte in the zygote, an additional mechanism may have evolved (Mayer et al. 2000). It has been suggested that the preponderance of antisense transcripts seen for imprinted genes that act to suppress the paternal allele is to replace the repressive activity of DNA methylation, and is therefore another weapon in the sexual conflict theory of imprinting.

Although it is widely held that DNA methylation is important to mammalian development (Eden and Cedar 1994; Meehan 2003), it is not without debate (Walsh and Bestor 1999). The vast majority of methylated DNA in the mammalian genome is found at transposable repetitive elements (Yoder et al. 1997). It has been proposed that this is the true function of DNA methylation, as host defense against transposition and that its use in imprinting and X-inactivation may be secondary and may not be important for development otherwise (Bestor 2000). This is still controversial and the fact that transpositions are still controlled during the stages of demethylation argues against this (Bird 1997). Others have suggested DNA methylation is necessary to control background transcriptional noise (Bird 1997) or to modify the regulatory effects of transposed elements indirectly (Martienssen 1998). Whatever the evolutionary function of DNA methylation, it clearly has a major role in imprinted gene regulation.

Histone modifications

Nucleosomes are the basic monomeric unit of chromatin. With an amazing capacity to compact two meters of DNA into a eukaryotic nucleus, it was easy to assume the nucleosome's function was one of simple compression. Of course it is now known that histones plays an important role in gene regulation (Strahl and Allis 2000; Felsenfeld and Groudine 2003). The basic structure of the nucleosome consists of two each of the four core histone subunits H4, H3, H2A and H2B, with approximately 146 bp of DNA coiled around a basic groove in the complex. As a polymer, nucleosomes are central to higher order chromatin organization (Figure 1-7), especially in the 11 nm and 30 nm configurations where the unstructured histone tail domains are exposed and accessible to modification.

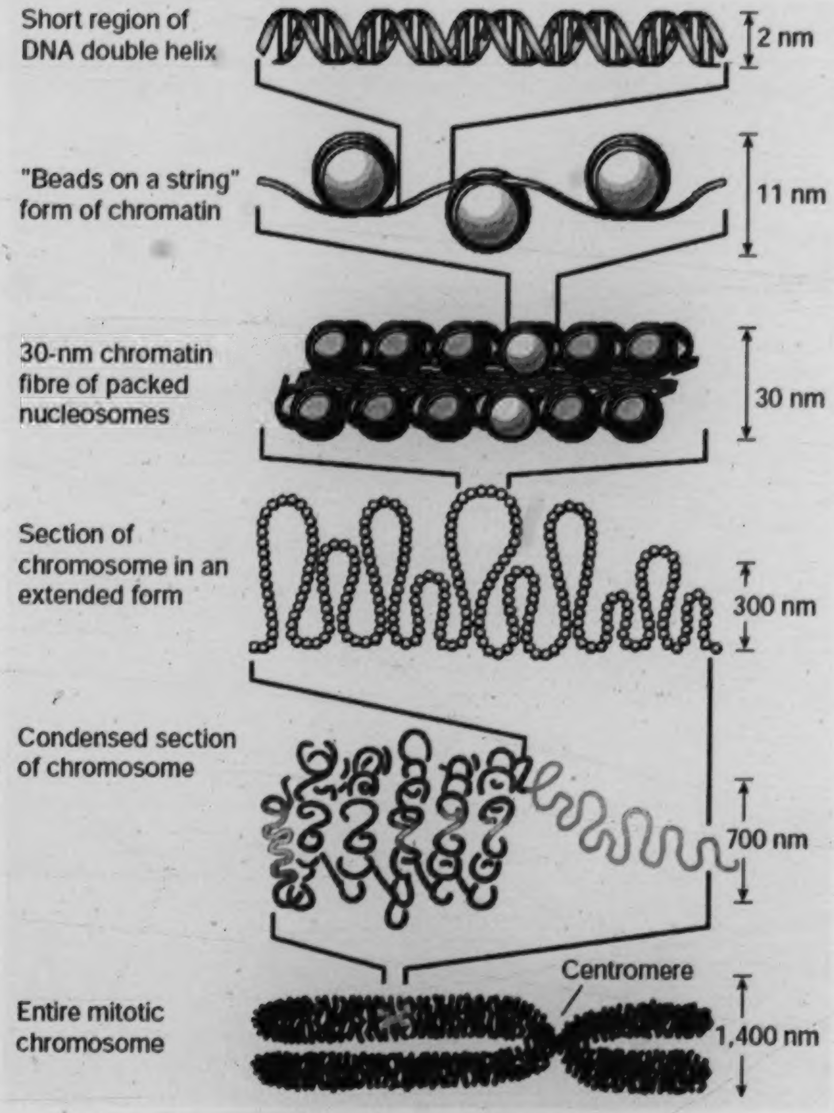

Short region of DNA double helix
2 nm
"Beads on a string" form of chromatin
11 nm
30-nm chromatin fibre of packed nucleosomes
30 nm
Section of chromosome in an extended form
300 nm
Condensed section of chromosome
700 nm
Centromere
Entire mitotic chromosome
1,400 nm

Figure 1-7. Levels of chromatin compaction.
Sequential compaction of DNA allows compression as well as regulation at different levels. Reproduced from Figure 1A of (Felsenfeld and Groudine 2003).

Despite its status as darling of the epigenetics field in recent years, histone modification as it pertains to gene regulation is not a new idea. Histone acetylation was associated with gene activation and transcription almost forty years ago (Pogo et al. 1966), but of course without the detailed *in vivo* characterization that was made possible by chromatin immunoprecipitation (ChIP) (Kuo and Allis 1999). Perhaps more crucially, antibody reagents capable of recognizing specifically modified histones in ChIP were developed (Hebbes et al. 1988). The ChIP assay represented for the first time a method by which protein-DNA interactions *in vivo* can be studied for any endogenous locus in a sequence-specific manner. Since then, the field advanced as fast as new antibody specificities could be produced and characterized. A new nomenclature evolved to deal with this new language (Turner 2005). For simplicity, the rest of this text will use this nomenclature convention for discussion involving histone modifications (Table 1-1). With each new report detailing association between a chromatin-templated biological process and a specific histone modification, it became clear this represented a novel level of information in gene regulation (Figure 1-8, Table 1-2). The histone code hypothesis was proposed, which suggests that combinations of multiple covalent modifications on histones specify unique biological function (Strahl and Allis 2000; Grant 2001; Iizuka and Smith 2003; Peterson and Laniel 2004; Dion et al. 2005).

Modifying group	Amino acid(s) modified	Level of modification	Abbreviation for modification	Examples of modified residues
Acetyl-	Lysine	mono-	ac	H3K9ac
Methyl-	Arginine	mono-	me1	H3R17me1
	Arginine	di-, symmetrical	me2s	H3R2me2s
	Arginine	di-, asymmetrical	me2a	H3R17me2a
	Lysine	mono-	me1	H3K4me1
	Lysine	di-	me2	H3K4me2
	Lysine	tri-	me3	H3K4me3
Phosphoryl-	Serine or threonine	mono-	ph	H3S10ph
Ubiquityl-	Lysine	mono-	ub1	H2BK123ub1
SUMOyl-	Lysine	mono-	su	H4K5su
ADP ribosyl-	Glutamate	mono-	ar1	H2BE2ar1
	Glutamate	poly-	arn	H2BE2arn

Table 1-1. Brno nomenclature for histone modifications. Abbreviation scheme for histone modifications, with subunit first, then residue, and modification. Named after Brno, Czech Republic, where the nomenclature was proposed. Reproduced from Table 1 of (Turner 2005).

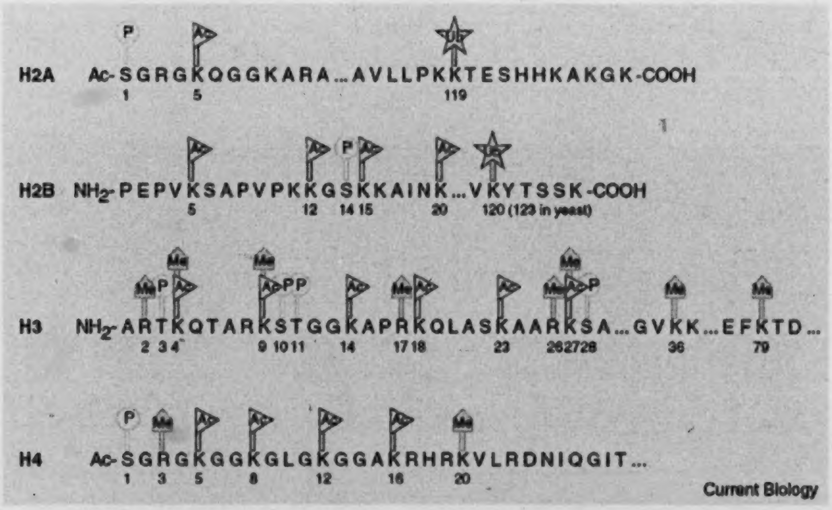
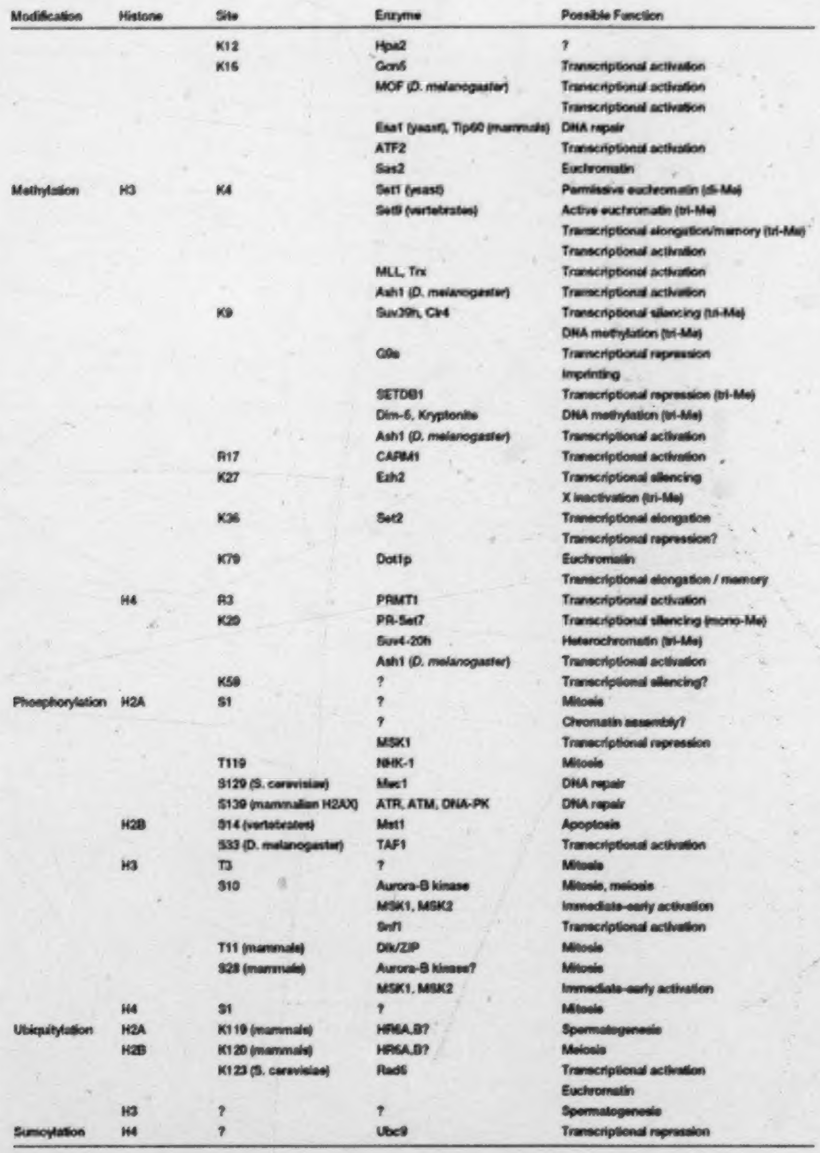

H2A
Ac-SGRGKQGGKARA...AVLLPKKTESHHKAKGK-COOH
1 5 119
H2B
NH2-PEPVKSAPVPKKGSKKAINK...VKYTSSK-COOH
5 12 14 15 20 120 (123 in yeast)
H3
NH2-ARTKQTARKSTGGKAPRKQLASKAARKSA...GVKK...EFKTD...
2 3 4 9 10 11 14 17 18 23 26 27 28 36 79
H4
Ac-SGRGKGGKGLGKGGAKRHRKVLRDNIQGIT...
1 3 5 8 12 16 20
P
Ac
Ub
Me
Current Biology

Figure 1-8. Covalent modifications of canonical histone subunit tails. All four histone tails can carry covalent modifications. Ac = acetylation, P = phosphorylation, Me = methylation, Ub = ubiquitination. Reproduced from Figure 1 of (Peterson and Laniel 2004).

Modification	Histone	Site	Enzyme	Possible function
Acetylation	H2A	K4 (*S. cerevisiae*)	Esa1	Transcriptional activation
		K5 (mammals)	Tip60	Transcriptional activation
			p300/CBP	Transcriptional activation
		K7 (*S. cerevisiae*)	Hat1	?
			Esa1	Transcriptional activation
	H2B	K5	ATF2	Transcriptional activation
		K11 (*S. cerevisiae*)	Gcn5	Transcriptional activation
		K12 (mammals)	p300/CBP	Transcriptional activation
			ATF2	Transcriptional activation
		K16 (*S. cerevisiae*)	Gcn5	Transcriptional activation
			Esa1	
		K15 (mammals)	p300/CBP	
			ATF2	Transcriptional activation
		K20	p300	Transcriptional activation
	H3	K4	Esa1	Transcriptional activation
			Hpa2	?
		K9	?	Histone deposition
			Gcn5	Transcriptional activation
			SRC-1	Transcriptional activation
		K14	Gcn5, PCAF	Transcriptional activation
			Esa1, Tip60	Transcriptional activation
				DNA repair
			SRC-1	Transcriptional activation
			Elp3	Transcription elongation
			Hpa2	?
			hTFIIIC90	RNA polymerase III transcription
			TAF1	RNA polymerase II transcription
			Sas2	Euchromatin?
			Sas3	Transcriptional activation/elongation?
			p300	Transcriptional activation
		K18	Gcn5 (SAGA/STAGA complex)	Transcriptional activation
				DNA repair
			p300, CBP	DNA replication
				Transcriptional activation
		K23	Gcn5 (SAGA/STAGA complex)	Transcriptional activation
			Sas3	DNA repair
			p300, CBP	Transcriptional activation/elongation?
				Transcriptional activation
		K27	Gcn5	Transcriptional activation
	H4	K5	Hat1	Histone deposition
			Esa1, Tip60	Transcriptional activation
				DNA repair
			ATF2	Transcriptional activation
			Hpa2	?
			p300	Transcriptional activation
		K8	Gcn5, PCAF	Transcriptional activation
			Esa1, Tip60	Transcriptional activation
				DNA repair
			ATF2	Transcriptional activation
			Elp3	Transcription elongation
			p300	Transcriptional activation
		K12	Hat1	Histone deposition
				Telomeric silencing
			Esa1, Tip60	Transcriptional activation
				DNA repair

Table 1-2. Summary of modifications.
Table of modifications, their modifiers and functions. Reproduced from (Peterson and Laniel 2004).



Modification	Histone	Site	Enzyme	Possible Function
		K12	Hpa2	?
		K16	Gcn5	Transcriptional activation
			MOF (*D. melanogaster*)	Transcriptional activation
				Transcriptional activation
			Esa1 (yeast), Tip60 (mammals)	DNA repair
			ATF2	Transcriptional activation
			Sas2	Euchromatin
Methylation	H3	K4	Set1 (yeast)	Permissive euchromatin (di-Me)
			Set9 (vertebrates)	Active euchromatin (tri-Me)
				Transcriptional elongation/memory (tri-Me)
				Transcriptional activation
			MLL, Trx	Transcriptional activation
			Ash1 (*D. melanogaster*)	Transcriptional activation
		K9	Suv39h, Clr4	Transcriptional silencing (tri-Me)
				DNA methylation (tri-Me)
			G9a	Transcriptional repression
				Imprinting
			SETDB1	Transcriptional repression (tri-Me)
			Dim-5, Kryptonite	DNA methylation (tri-Me)
			Ash1 (*D. melanogaster*)	Transcriptional activation
		R17	CARM1	Transcriptional activation
		K27	Ezh2	Transcriptional silencing
				X inactivation (tri-Me)
		K36	Set2	Transcriptional elongation
				Transcriptional repression?
		K79	Dot1p	Euchromatin
				Transcriptional elongation / memory
	H4	R3	PRMT1	Transcriptional activation
		K20	PR-Set7	Transcriptional silencing (mono-Me)
			Suv4-20h	Heterochromatin (tri-Me)
			Ash1 (*D. melanogaster*)	Transcriptional activation
		K59	?	Transcriptional silencing?
Phosphorylation	H2A	S1	?	Mitosis
			?	Chromatin assembly?
			MSK1	Transcriptional repression
		T119	NHK-1	Mitosis
		S129 (*S. cerevisiae*)	Mec1	DNA repair
		S139 (mammalian H2AX)	ATR, ATM, DNA-PK	DNA repair
	H2B	S14 (vertebrates)	Mst1	Apoptosis
		S33 (D. melanogaster)	TAF1	Transcriptional activation
	H3	T3	?	Mitosis
		S10	Aurora-B kinase	Mitosis, meiosis
			MSK1, MSK2	Immediate-early activation
			Snf1	Transcriptional activation
		T11 (mammals)	Dlk/ZIP	Mitosis
		S28 (mammals)	Aurora-B kinase?	Mitosis
			MSK1, MSK2	Immediate-early activation
	H4	S1	?	Mitosis
Ubiquitylation	H2A	K119 (mammals)	HR6A,B?	Spermatogenesis
	H2B	K120 (mammals)	HR6A,B?	Meiosis
		K123 (*S. cerevisiae*)	Rad6	Transcriptional activation
				Euchromatin
	H3	?	?	Spermatogenesis
Sumoylation	H4	?	Ubc9	Transcriptional repression

Table 1-2, continued. Summary of modifications.
Table of modifications, their modifiers and functions. Reproduced from (Peterson and Laniel 2004).

Histones are also one of very few candidates for carrying epigenetic information. It is clear how DNA methylation is transmitted past DNA replication, and there are models to suggest how histone modifications can be transmitted. Most importantly, it has been shown that nucleosomes are retained on daughter strands past the replication fork (Figure 1-9 A) (Bonne-Andrea et al. 1990), and it is thought that parental histones are distributed equally to newly synthesized DNA with the addition of new nucleosomes (Krude 1999; Lucchini et al. 2001). This dilution of parental histones and their covalent modifications is strikingly analogous to the hemi-methylated state of newly synthesized DNA. Spread and propagation of the modification state of parental nucleosomes to new nucleosomes can then accomplish maintenance of these states through cell divisions. This is still controversial however as some sites of modification thought to be important in gene regulation span only several histones, raising the doubt that faithful distribution to daughter strands is possible (Henikoff et al. 2004). A model by which histone halves are split off to daughter strands, while attractive in terms of epigenetic transmission, has very little experimental support (Figure 1-9 B). These states are obviously propagated though, and in some cases without DNA methylation, so it is likely a matter of when, instead of if, these mechanisms will be found. In fact, an alternate model of epigenetic inheritance involving variant histones will be discussed later (Figure 1-10).

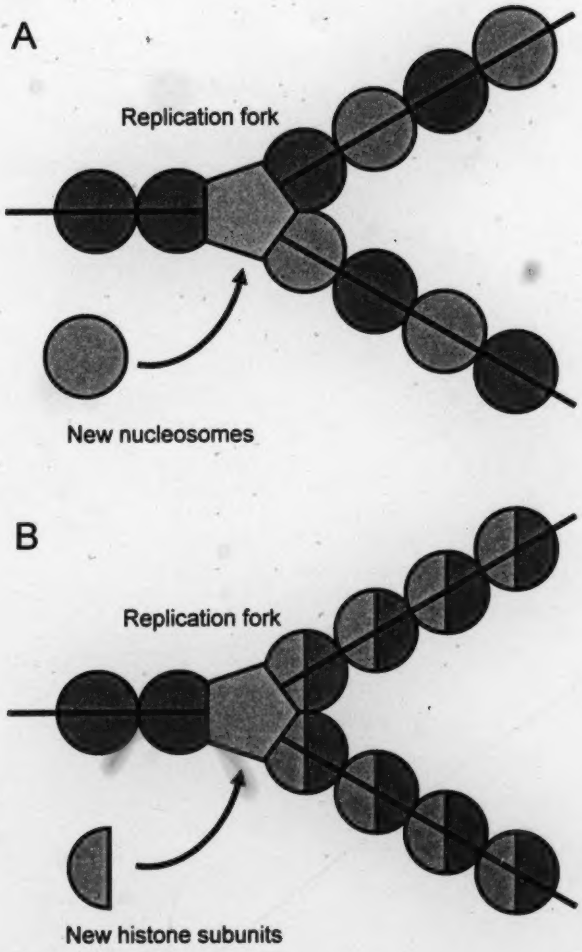

A
Replication fork
New nucleosomes
B
Replication fork
New histone subunits

Figure 1-9. Models of histone conservation during replication.
(A) Nondispersive model where old nucleosomes (dark brown) are segregated randomly to leading and lagging strands past the replication fork (green pentagon) with new nucleosomes (light brown). (B) Semi-conservative model where new subunits are incorporated with old subunits. In this model, it is one pair of H3-H4 and H2A-H2B that is conserved with new subunits.

As with DNA methylation, there are proteins able to read, write and erase the information on histones. But unlike the biochemical simplicity of DNA methylation, histone modification is highly variable and requires families of enzymes, and variants of domains to carry out downstream biological directives. Many of these enzymes have previously been found to be transcriptional co-activators or co-repressors (Grant 2001). Histone acetylation has been most well characterized (Grunstein 1997). It is associated with gene activity as well as potentiation of transcription (Hebbes et al. 1988). Families of histone acetyltransferases (HATs) and histone deacetylases (HDACs) catalyze this modification on lysine residues on all four subunits of the nucleosome. Many HATs and HDACs have been characterized, some with preferences for certain residues. Two types of HATs exist: nuclear type-A, involved in gene regulation, and cytoplasmic type-B, involved in pre-acetylation of newly synthesized histones for assembly. GCN5 is one of the first HATs characterized in yeast and it is required for activation of many genes (Kuo et al. 1996). Other HATs such as CBP and p300 have been shown to be involved in gene activation in response to cellular signals (Chakravarti et al. 1996). They also do not act alone, but participate in large complexes that are able to target their activities. Some are also able to acetylate other proteins, often contributing to gene activation. Similarly, HDACs are part of large repressive complexes and are recruited to genes through complex interactions (Grunstein 1997). While the acetylation of histones can bring about changes in chromatin structure by neutralizing the positive charge that facilitates interaction with DNA, acetylation can also act through signal transduction by creating binding sites for the bromodomain (de la Cruz et al. 2005; Dion et al. 2005). Analogous to the function of the

MBD in the recognition of DNA methylation, bromodomains bind acetylated histones and are found in HATs, chromatin remodeling enzymes and HMTs, indicating their importance in chromatin and gene regulation. The observation that HATs contain bromodomains may be a significant indicator of how these modifications can be amplified or propagated to nearby residues, adjacent histone subunit tails, as well as neighboring nucleosomes.

In contrast to the harmonious picture of how histone acetylation functions in transcriptional regulation, histone phosphorylation seems to be involved in many different pathways. Phosphorylation was first linked to mitosis (Hsu et al. 2000; Nowak and Corces 2004), but has also been associated with transcriptional control (Peterson and Laniel 2004). Again, this may be through changes in charge of the nucleosomal particle or binding of specific factors, or most intriguingly, through interactions with other modifications, which will be discussed later. Recently, H2BS14ph has been shown to be involved in apoptosis by a specific kinase Mst1 (Ahn et al. 2005). Phosphorylation of histone variant H2A.X during damage has been shown to have an important role in DNA double strand break repair by INO80 complex recruitment (Morrison et al. 2004). Covalent addition of ubiquitin and SUMO groups has also been observed on histones and is linked to various chromatin functions. Unlike the poly-ubiquitination that is associated with protein turnover, histones are often mono-ubiquitinated, which may be related to mitosis, meiosis (Robzyk et al. 2000), transcription (Davie and Murphy 1994), and most importantly, to spermatogenesis (Jason et al. 2002). Sumoylation has been discovered recently and is associated with repression (Shiio and Eisenman 2003). ADP-ribosylation of histones is not well understood, but there is evidence this modification of the linker histone H1 may be involved in memory formation (Cohen-Armon et al. 2004). While these processes are fascinating in their own right, this discussion will be restricted to modifications most relevant to transcriptional regulation.

Histone methylation has taken the limelight in recent years. While most histone acetylation states are associated with transcriptional activation, histone methylation can serve many roles, including activation and repression of gene activity (Peterson and Laniel 2004). Histone tails contain arginine (R) and lysine (K) residues that can be methylated. Furthermore, R and K residues can accommodate up to two and three methyl groups respectively, to elaborate the signal. In addition, dimethylation of R can be in two steric variants (Bannister et al. 2002). It is also believed that histone methylation has a longer half-life than acetylation or phosphorylation, making it a better candidate for long term memory of gene activity (Bannister et al. 2002). The first example was the characterization of CARM1 that is able to methylate H3 R residues and participates in the activation of genes under control of nuclear hormone receptors (Chen et al. 1999). It was soon found that H3K4me was a well conserved modification that was involved in gene activity (Strahl et al. 1999). Different levels of methylation at this lysine residue also correlated with different levels of activity; in this case, H3K4me2 and H3K4me3 with euchromatin and transcriptionally active genes, respectively (Santos-Rosa et al. 2002). H3K79me has been associated with telomeric silencing in yeast (Lacoste et al. 2002) and gene activation in yeast and mammals (Ng et al. 2003). In contrast, methylation of H3K9 is involved in heterochromatin assembly (Nakayama et al. 2001). This modification can be made by SUV39H1, the human homolog of a *Drosophila* gene long known to be involved in gene regulation and position effects (Rea et al. 2000). SUV39H1 and other histone HMTs that act on lysines contain an evolutionarily conserved SET (Su(var), E(z), trithorax) domain, found in many proteins involved in transcriptional regulation (Xiao et al. 2003). While SUV39H1 is involved in the H3K9me of heterochromatin, another HMT, G9a is involved in the H3K9 and K27 methylation at euchromatic sites, necessary in transcriptional repression of developmental genes essential in embryogenesis (Tachibana et al. 2002; Roopra et al. 2004). Interestingly, G9a is essential in imprinting of the PWS/AS IC (Xin et al. 2003). Similar to bromodomain bearing proteins that carry out downstream effects of histone acetylation, specific proteins recognize histone methylation. H3K4me, a mark of activity, is able to disrupt binding of the NuRD repressor complex (Nishioka et al. 2002; Zegerman et al. 2002). H3K9me3, a mark of silencing, can be bound by the chromodomain of HP1, a major facultative heterochromatic protein (Bannister et al. 2001; Lachner et al. 2001). H3K27me however, is bound by polycomb group proteins (PcG) in regulation of developmental gene clusters (Kirmizis et al. 2004). H3K9me and H3K27me represent extremely similar epitopes, as evidenced by some antibodies that cannot distinguish the two, but they are uniquely identified by HP1 and PcG, illustrating the biological specificity of these marks (Fischle et al. 2003c). Subtle adjustment of H3K9 and K27 methylation levels are associated with facultative and constitutive heterochromatin with H3K9me3 and H3K27me1 at pericentric heterochromatin and H3K9me1/2 and H3K27me1/2/3 at silent euchromatin. With the number of residues and variation of methylation of each residue, a complex code can be elucidated from just histone methylation (Craig 2005). Recent advances have also shown that the histone methylation mark can truly be regulated by erasure instead of the histone replacement during replication (Shi et al. 2004). Much like HDACs, LSD1 was identified as a co-repressor that specifically demethylates the H3K4 residue. It remains to be seen whether other histone methylation marks such as H3K9 and H3K27 can also be demethylated by a similar enzyme, but discovery of LSD1 will facilitate this search. It is also important to remember that while many modifying enzymes have been termed histone-specific, their substrates may not be so restricted, and HATs, HDACs, and HMTs may modify other proteins, often also involved in gene regulation (Chen et al. 1999; Robzyk et al. 2000; Girdwood et al. 2003).

Another way to reverse the effects of histone methylation or any other modification is the replacement of that histone, either during DNA replication or by replication independent means (Ahmad and Henikoff 2002). While canonical histone H3 subunits are only synthesized and incorporated during replication, variants can be added at other times, and by special complexes. H3.3 is a special H3 replacement variant that differs by only four amino acids and is incorporated into heavily transcribed euchromatin (Tagami et al. 2004). Even variants are modified as canonical histones are. H3.3 is often enriched in active modifications (McKittrick et al. 2004). It has been proposed that epigenetic inheritance of chromatin states can be achieved through the use of H3.3 (Henikoff et al. 2004). An actively transcribed region would be enriched in H3.3 which, after dilution to daughter strands during replication would direct the transcription of the locus in the daughter cell, thereby enriching the region again for H3.3 (Figure 1-10). In this scheme, it is transcription that takes on the role of maintenance, versus amplification of histone modifications having this role in the previous model of transmission of histone states. Note that of course, these are not mutually exclusive models. It is also suggested that this is how the activation states of certain regulatory elements are transmitted through intergenic transcription (Rank et al. 2002). It should be noted here, that this model is distinct from antisense regulation which inactivates genes by transcription in *cis*.

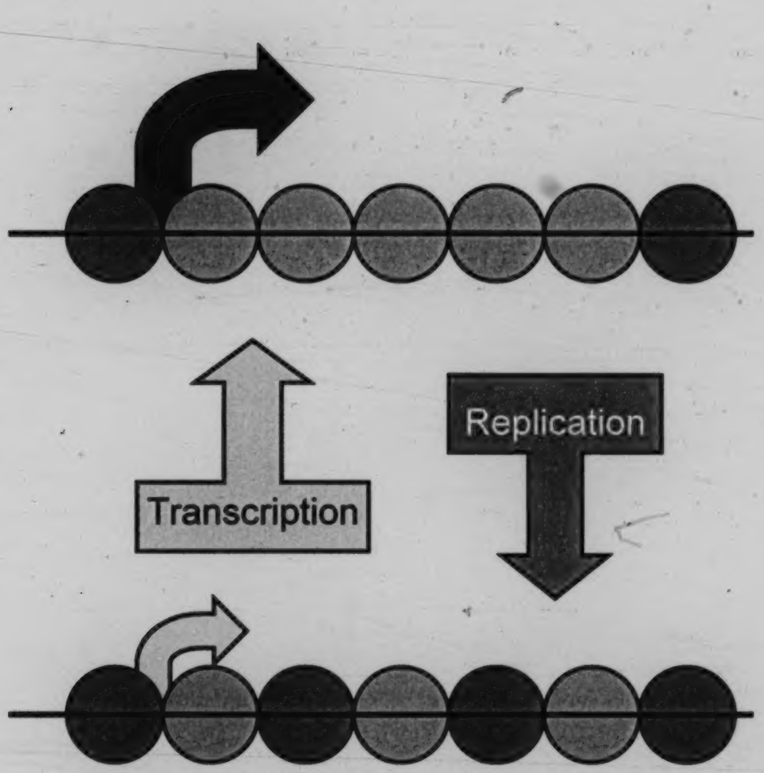
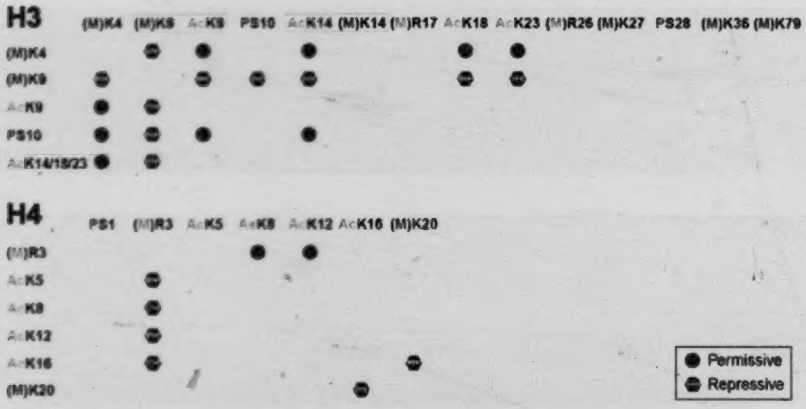


Figure 1-10. Mechanism of transmission of H3.3 states through replication.

Transcription causes local replacement with variant H3.3. Replication would lead to a dilution of H3.3, but still leads to transcription of the locus, causing more replacement with H3.3, completing the cycle.

Adding yet another level of complexity is the central tenet of the histone code hypothesis: that different modifications act in sequence or in combinations to bring about unique biological properties (Strahl and Allis 2000). The most obvious example is the physical occupancy of a residue by a modification, thereby preventing further reactions. For example, the methylation of H3K9 would preclude acetylation at the same site. More commonly, examples have shown interactions between different residues (Figure 1-11) (Fischle et al. 2003b). H3S10ph has been shown to be a prerequisite for and to promote H3K14ac (Cheung et al. 2000; Lo et al. 2000). H3S10ph however, can be inhibited by methylation of the adjacent residue to H3K9me (Rea et al. 2000). To bring these interactions full circle, H3K9me depends on H3K14 deacetylation (Nakayama et al. 2001). This interactive and synergistic cycle may specify an ON and OFF state (Figure 1-12) (Berger 2001); other such "binary switches" have been characterized (Fischle et al. 2003a). Interactions can also be extremely specific: for example, H3K4me by Set9 inhibits Suv39h1 but not G9a mediated H3K9me (Nishioka et al. 2002). Modification on different tails can also affect each other. The best example is the unidirectional requirement of H2BK123ub for H3K4me, where a mutation of H2B K123R abolishes H3K4me but H3 K4R does not affect H2BK123ub levels (Sun and Allis 2002). One question comes to mind in consideration of these complex networks. Why must there be so many marks seemingly contributing to the same outcome? Perhaps slight adjustment of the code allows finer tuning than the simplicity of DNA methylation allows. The fluidity of histone modifications and the number of modifying enzymes in many pathways may also represent a point of integration of numerous cellular signals in gene regulation.



H3	(M)K4	(M)K9	AcK9	PS10	AcK14	(M)K14	(M)R17	AcK18	AcK23	(M)R26	(M)K27	PS28	(M)K36	(M)K79
(M)K4		⊖	●		●			●	●					
(M)K9	⊖		⊖	⊖	⊖			⊖	⊖					
AcK9	●	⊖												
PS10	●	⊖	●		●									
AcK14/18/23	●	⊖												

H4	PS1	(M)R3	AcK5	AcK8	AcK12	AcK16	(M)K20
(M)R3				●	●		
AcK5		⊖					
AcK8		⊖					
AcK12		⊖					
AcK16		⊖					⊖
(M)K20						⊖	

● Permissive
⊖ Repressive

Figure 1-11. Examples of interactions between modifications.
Examples of permissive and inhibitory interactions between modifications on the histone tail on further modification of the same tail. Reproduced from Figure 1 of (Fischle et al. 2003b)

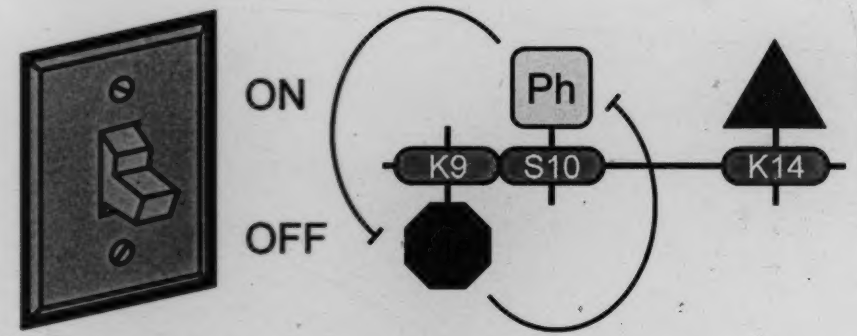


Figure 1-12. Binary switch theory.
Example of a binary switch of gene transcription, where one set of modifications synergistically has the opposite action of different set of modifications. Certain individual modifications may also have repressive effects on others.

Epigenetic interaction

An epigenetic hierarchy has often been sought, most importantly, between DNA methylation and histone modifications. Which is upstream in the pathway of gene regulation? As previously described, methylated DNA can recruit histone deacetylase activities through MBD containing proteins and complexes involved in repression (Bird and Wolffe 1999). Some proteins have also been found to contain a SET domain and a putative MBD, raising the possibility that DNA methylation can also direct histone methylation (Figure 1-13) (Zhang and Reinberg 2001). With the well established role of DNA methylation in gene regulation and its clear mode of inheritance, it was easy to speculate that it was higher up in the epigenetic hierarchy. Contrary to this, it was shown that in *Neurospora*, DNA methylation requires H3K9me. It was discovered that a mutation that causes decrease in DNA methylation (*dim-5*) was a HMT (Tamaru and Selker 2001). A similar situation was found in plants which carry CpNpG methylation as well as CpG methylation. Mutations in a plant HMT *kyp* abolished CpNpG methylation and H3K9me, but mutation of the CpNpG methylase *cmt3* only affected DNA methylation and not H3K9me (Johnson et al. 2002). It was later found that CMT3 bound, and required both H3K9me and H3K27me together to direct CpNpG methylation, showing the mechanism by which histone modifications can direct DNA methylation and adding to the histone code for silencing (Lindroth et al. 2004). Interestingly, histone modifications are more conserved across eukaryotes than DNA methylation, which is absent or negligible in organisms such as *S. pombe* and *Drosophila* (Nakayama et al. 2001). It has been proposed that DNA methylation is a more recent addition to the epigenetic schema to allow elaboration and perhaps more stable marking of expression (Rice and Allis 2001), or as mentioned earlier, as a protective agent against genomic parasites (Bestor 2000). This is supported by the observation that the *cmt3* and *kyp* mutants show reactivation of certain retrotransposons in *Arabidopsis* (Bartee et al. 2001).

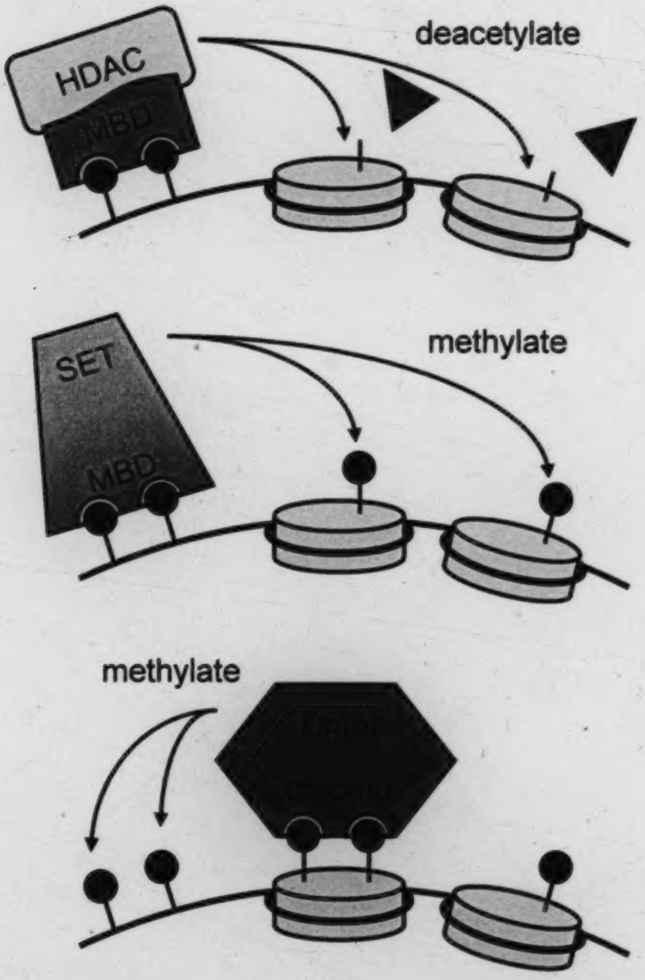

HDAC
MBD
deacetylate
SET
MBD
methylate
methylate
Dnmt
Chromo

Figure 1-13. Interaction between DNA methylation and histone modification.

Examples of how DNA methylation and histone modifications can interact with each other. Repressive MBD proteins can bind methylated DNA (lollipop) and recruit HDAC activities to deacetylate histones. More directly, a protein can have both a MBD, as well as a histone modification domain such as a SET domain to methylate histones. Alternatively, a protein can bind methylated histones through a chromodomain and methylate DNA.

There are also important connections between gene regulation, histone modifications, DNA methylation and RNAi (RNA interference) (Matzke and Birchler 2005). RNAi is involved in aspects of transcriptional control at the RNA and chromatin level through the use of short homologous RNA molecules. Based on studies in plants, it has been found to be able to destroy specific mRNA and cause epigenetic alterations of DNA targets, but the existence of this pathway in mammals is still in question. As my work does not deal with RNAi and the links between its mechanisms and those proposed for antisense regulation of mammalian imprinted genes and X-inactivation are tenuous, this is outside the scope of this introduction.

Epigenetic mechanisms in genomic imprinting

The epigenetic hierarchy is especially pertinent to imprinting and X-inactivation. What comes first and what gets transmitted is central to the question of how two alleles in the same cell can maintain differential expression and epigenetic states independently of each other. With the mechanisms of transmission of histone modifications uncharacterized, DNA methylation is still the prime candidate for the imprinted mark, especially between generations, with the observation that ICs can carry differential methylation from the gametes (Soejima and Wagstaff 2005). However, the histone variant CENP-A is not replaced by protamines during spermatogenesis and is thought to be the mechanism by which centromeres are epigenetically transmitted (Henikoff et al. 2004), opening the possibility that a small subset of histones or their variants can also be retained. Histone variants are also correlated with the active and inactive X's (macroH2A and H2A-Bbd) (Chadwick and Willard 2001; Chadwick and Willard 2002). The relationship so far between DNA methylation, histone modification and imprinting has varied by report, and no unifying theory exists. It is known that in *Dnmt1* null embryos, imprinting is perturbed for some but not other genes (Caspary et al. 1998; Xin et al. 2003). Acetylation was affected for H4 but not H3 in treatments with a DNA methylation inhibitor at *SNRPN* in human cell culture and resulted in reactivation of the silent maternal allele (Saitoh and Wada 2000). In mouse cell culture however, neither treatment with a HDAC inhibitor nor with the DNA methylation inhibitor was able to reactivate the silent allele of *Snrpn*, although some other imprinted genes were reactivated, and in some cases, in a heritable fashion after withdrawal of inhibitors (El Kharroubi et al. 2001). In *G9a*-null embryos, there is a loss of imprinting for *Snrpn* (Xin et al. 2003). This is in contrast to the *Dnmt1* null that exhibits loss of differential DNA methylation at the IC, but retains differential H3K9me and imprinting. H3K9me is also one of the earliest events in X-inactivation after Xist coating (Heard et al. 2001; Mermoud et al. 2002). In a report studying which of DNA methylation, histone modifications, or antisense transcription was more correlated with tissue specific imprinting of *IGF2R/Igf2R* in human and mouse, it was found that H3K4me and H3K9me were most consistent with imprinted expression (Vu et al. 2004). Imprinting of the *Kcnq1* imprinted domain in the placenta, which does not involve differential DNA methylation, is instead associated with H3K27me3, H3K9me2, and the PcG complex Eed-Ezh2 (Umlauf et al. 2004). The same PcG complex and H3K27me is also involved in X-inactivation (Plath et al. 2003), further implicating the evolutionary commonality between imprinting and X-inactivation (Lee 2003). Therefore, it is obvious that genomic imprinting and X-inactivation employ multiple epigenetic mechanisms in mono-allelic gene regulation, but the players at various loci, tissues and species differ. Further examination of mechanisms involved will be instrumental in deciphering the language necessary for imprinting.

CTCF Biology

CTCF, the mammalian insulator protein

CTCF was first named for its ability to bind a CCCTC DNA sequence (Lobanenkov et al. 1990). But in fact, it is able to bind an array of different sequences in many species (Figure 1-14) (Ohlsson et al. 2001). It is highly conserved across vertebrates, and contains 11 zinc finger (ZF) domains, which gives it the ability to bind divergent sequences through combinatorial use of these domains (Filippova et al. 1996). Unlike other zinc finger proteins, the ZFs in CTCF are able to interact with proteins as well as DNA (Ohlsson et al. 2001). CTCF is able to activate or inhibit transcription in different situations depending on interactions with other proteins (Baniahmad et al. 1990; Awad et al. 1999). It is also able to recruit HDACs, linking it to histone modification (Lutz et al. 2000). Perhaps its most important function is at mammalian insulators. While the mechanism of insulator function is unknown (Bulger and Groudine 1999), it is thought that most if not all mammalian insulators bind CTCF (Bell et al. 1999). As mentioned earlier, CTCF function at an allelic insulator has a pivotal role in *Igf2/H19* imprinting (Bell and Felsenfeld 2000; Hark et al. 2000).

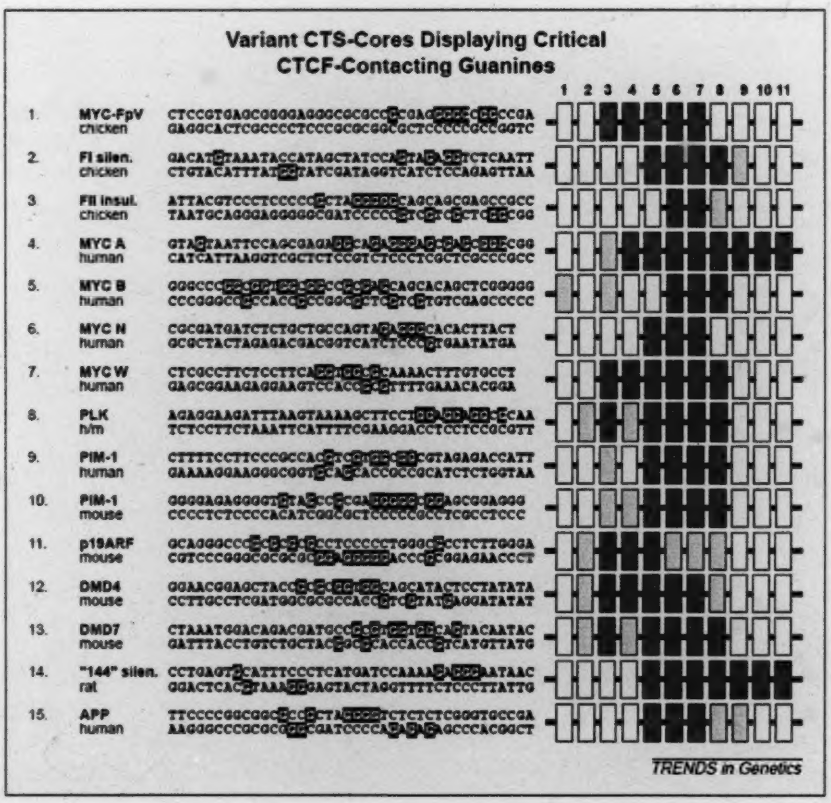

Variant CTS-Cores Displaying Critical CTCF-Contacting Guanines
1 2 3 4 5 6 7 8 9 10 11
1. MYC-FpV chicken
2. FI silen. chicken
3. FII insul. chicken
4. MYC A human
5. MYC B human
6. MYC N human
7. MYC W human
8. PLK h/m
9. PIM-1 human
10. PIM-1 mouse
11. p19ARF mouse
12. DMD4 mouse
13. DMD7 mouse
14. "144" silen. rat
15. APP human
TRENDS in Genetics

Figure 1-14. Variation in ZF domain usage in DNA recognition. Chart showing guanine residues important in a subset of CTCF binding sites on the left. On the right, are the corresponding ZF domains used to contact those guanines, with white being most important and red dispensable in deleted CTCF constructs (pink being incomplete loss of binding when deleted). DMD4 and DMD7 are from the *Igf2/H19* ICR. Reproduced from Figure 2 of (Ohlsson et al. 2001).

One of the most well characterized CTCF binding insulators is the locus control region (LCR) of the β-globin genes (Bell et al. 1999). The β-globin LCR in many ways was at the forefront of the field of epigenetics. It is able to confer tissue and temporal-specific regulation to the individual globin genes and contains numerous DNase I hypersensitive sites, a hallmark of regulatory elements (Li et al. 1999). One of these sites (HS4) is able to block enhancers and prevent position effects, the latter of which is dependent on CTCF (Recillas-Targa et al. 2002). This site is also associated with the nuclear matrix, which is also dependent on CTCF (Yusufzai and Felsenfeld 2004). The CTCF binding site is also responsible for blocking of the spread of histone acetylation (Zhao and Dean 2004). Conserved CTCF sites flank the entire β-globin domain and may serve as delimiters of the domain from outside influence (Saitoh et al. 2000). Before the looping structures at the imprinted *Igf2/H19* locus were discovered (Murrell et al. 2004), a similar looping structure was found for the LCR and the transcribed globin genes (Carter et al. 2002). The β-globin LCR has shown the way in the study of imprinting and other epigenetic phenomenon, and parallels continue to come to light. In a way, ICs can be thought of as LCRs that confer allele instead of tissue-specific regulation of distant genes.

From insulator to "imprintor"

The discovery that CTCF not only bound at the *Igf2/H19* DMR, but did so depending on methylation, gave important insights into CTCF function (Bell and Felsenfeld 2000; Hark et al. 2000). This became the first chromatin insulator controlled by DNA methylation, linking yet another epigenetic mark to CTCF other than HDAC recruitment. This link between DNA methylation and CTCF binding is especially relevant to cancer, where genome-wide hypomethylation may cause spurious CTCF binding and cause dysregulation of many genes (Mukhopadhyay et al. 2004). In fact, mutations in cancers have been characterized in CTCF that change its binding specificity (Filippova et al. 2002). This discovery also identifies CTCF as a trans-acting factor for the regulation of imprinting, at least at *Igf2/H19*. It has been suggested, but not proven, that sites also exist at the PWS/AS IC (Ohta et al. 2001) and at the *Dlk1-Gtl2* locus (Takada et al. 2002). A role for CTCF has also been found for the regulation of X-inactivation (Chao et al. 2002).

The story of CTCF and imprinting takes an interesting turn with the cloning of a new paralog that shares the same ZF domains, but differs in C and N terminal domains (Loukinov et al. 2002). Named *BORIS* (brother of regulator of imprinting), its most striking characteristic is that it is expressed in a mutually exclusive manner to CTCF. While *BORIS* is expressed in spermatocytes, *CTCF* is expressed in spermatogonia, spermatids and all other somatic tissues. It is expected that BORIS has the same DNA binding spectrum as CTCF, but be able to carry out different biological activities at those sites. This and the fact that it is only in the testis during the erasure of DNA methylation in spermatogenesis, suggests that BORIS may have a very important role in imprinting, and more specifically, during the initial setup of imprints and differentially methylated domains. Other evidence indicates *BORIS* also has an important role in cancer (Klenova et al. 2002). While CTCF has so far not been observed in other imprinted domains, evidence suggests the future work on the *CTCF/BORIS* brothers will reveal their roles in many aspects of gene regulation, epigenetics, and perhaps a more general role in imprinting.

Objectives and Rationale

Much is known about the genetic function of the PWS/AS IC in determining the allelic identity for genes spread out over several Megabases (Nicholls and Knepper 2001). This involves many aspects of epigenetic control including DNA methylation, histone modifications and chromatin conformation. Regulation through an antisense mechanism involving a long paternal transcript originating from the IC region is an elegant explanation and has received much support (Runte et al. 2001). The story of imprinting in the PWS/AS cluster is far from complete however as the paternal expression of the transposed imprinted genes, *NDN*, *MAGEL2*, *MKRN3* and *Frat3* falls outside the scope of this antisense model. Therefore, I have sought to uncover the epigenetic mechanisms underlying imprinting of the transposed imprinted genes. My approach will seek to understand imprinting at different scales and how imprinted regulation intersects with tissue-specific regulation.

Focus will begin at the level of an individual imprinted gene. Previous work in our laboratory has already shown differential chromatin accessibility at *Ndn* (Hanel 2003). A fine scale analysis of the regulatory elements of *NDN* will further indicate accessibility of the two alleles. This will allow an examination of footprints for possible factors involved in allele specification as well as tissue-specific regulation in tissues that do and do not express *NDN*. Chapter 3 will follow my *in silico* analysis of the promoter, and collaborative work with Dr. Meredith Hanel on the *in vivo* footprinting assays in *NDN*.

With the advent of ChIP, covalent modifications of histones in *NDN* can also be studied, but can be done at a scale of the entire transcription unit as well as surrounding regions. This will allow examination of the chromatin context *NDN* is in on either allele, and in different tissue types. It is conceivable that different histone codes specify imprinted versus developmental regulation. This study will also allow comparisons to histone modifications at the IC. It has been poorly characterized whether epigenetic characteristics at the IC are due to the imprinting function of the IC or transcription of the IC-associated transcription unit. Study of *NDN*, which is well isolated from the IC and is tissue-specific in expression, will help answer this question. Chapter 4 will reveal differences in the histone code correlated with allele-specific and tissue-specific regulation of *NDN*, using ChIP on human patient cells.

From the single representative transposed imprinted gene, it will be of great importance to evaluate if the other transposed imprinted genes are similar in their epigenetic regulation, as well as if these are conserved. If there are commonalities, perhaps mechanisms exist to co-regulate these genes. With the proposed role of CTCF in imprinting and function in long range control elements such as LCRs and insulators, it will be of value to study its role at the PWS/AS locus. This will aid in the understanding of locus-wide control of the transposed imprinted genes and give clues as to how their regulation fits with genes under control of the antisense transcript, as well as other imprinted genes. Chapter 5 will show evidence of an alternate mechanism of the regulation of the transposed imprinted genes that may be in a similar thread to that of the *Igf2/H19* locus.

The experiments presented here will also culminate in a more complete model of the regulation of the transposed imprinted genes, from allele-specific large scale domain regulation through CTCF, to local chromatin changes in histone modifications and factor binding reflective of tissue-specific differences. This will be discussed in Chapter 6.

Chapter 2 ♦ Materials and Methods

Parts of this chapter have appeared in:

Hanel ML, **Lau JC**, Paradis I, Drouin R, Wevrick R (2005) Chromatin modification of the human imprinted *NDN* (*necdin*) gene detected by *in vivo* footprinting. J Cell Biochem 94(5):1046-57

Lau JC, Hanel ML, Wevrick R (2004) Tissue-specific and imprinted epigenetic modifications of the human *NDN* gene. Nucleic Acids Res 32:3376-3382

Lau JC, Wevrick R (Submitted) CTCF binds differentially methylated regions in the imprinted mouse Prader-Willi Syndrome locus.

Sequence analysis

Transcription factor binding prediction

Promoter sequences of *NDN/Ndn* from human, mouse and rat were compared using ClustalW (Thompson et al. 1994) at the European Bioinformatics Institute (http://www.ebi.ac.uk/clustalw/) and Pairwise BLAST (Tatusova and Madden 1999). Percent identity plots were performed with Microsoft Excel, using ClustalW data. Prediction of transcription factor binding sites was performed with Genomatrix MatInspector (Quandt et al. 1995). Positions -495 to +193 with respect to the start codon in human NDN, -553 to +222 in mouse *Ndn*, and -554 to +223 in rat Ndn were analyzed for sequence conservation and putative transcription factor binding sites.

CTCF binding cluster prediction

Genomic fragments from the February 2003 freeze of the annotated sequence from the UCSC Genome Bioinformatics Browser (http://genome.ucsc.edu/, chr7: 49276508-51906627, approximately 87 kb centromeric to *Snurf-Snrpn*, and 80 kb telomeric to *Frat3*) (Karolchik et al. 2003) were searched for the CTCF degenerate consensus CCGCNNGGNGGCAG (Chao et al. 2002) using pDRAW32 (http://www.acaclone.com/), reducing stringency to allow 3 additional mismatches to compensate for the broad binding abilities of CTCF. Regions containing four or more sites within 100 bp of each other were chosen for further analysis. Of these regions, only those with unique sequence amenable to PCR analysis were chosen. DNA from *Mus castaneus* and *Mus musculus* was sequenced to identify polymorphisms that were then used for allele-specific analysis. Direct sequencing of PCR products was performed using fluorescent dye terminator reactions and an ABI sequencer.

Tissues and cultured cell lines

The availability of human cell lines derived from PWS and AS patients with deletions of the 15q11-q13 region represented a unique reagent in which alleles can be studied in isolation. Control fibroblasts from the NIGMS Human Genetic Cell repository (GM00650), PWS fibroblasts (our laboratory number FB16, University of Miami Brain and Tissue Bank for Developmental Disorders #1889), and AS fibroblasts (FB17, 15q11-q13 deletion cell line, from Dr. A. Beaudet, Baylor College of Medicine) were grown in DMEM supplemented with 10% fetal bovine serum (FBS). Control lymphoblastoid cell lines (LCL) derived from primary blood lymphocytes (LCL10), PWS LCLs (LCL3, GM09024B, GM09133) and AS LCLs (LCL1/2, GM11515) were grown in RPMI supplemented with 15% FBS. Blood was collected from PWS and AS patients with fluorescence *in situ* hybridization-verified deletions and from control individuals, with informed consent.

Animal protocols were approved by the University of Alberta Health Sciences Animal Policy and Welfare Committee. Mouse strains *Mus musculus* (C57BL/6) and *Mus castaneus* (CAST) were obtained from the Jackson Laboratories, USA. We used polymorphisms between these two strains to distinguish parental alleles in F1 progeny. Animals were euthanized with euthanol (Sigma) and cervical dislocation. Tissues were processed for DNA, RNA, or ChIP immediately.

DNA extraction and PCR

Genomic DNA was extracted from tissue culture cells by proteinase K/SDS digestion, phenol/chloroform extraction and ethanol precipitation (Ausubel et al. 1993). Blood was collected in sodium EDTA tubes and erythrocytes lysed in two successive washes with 4 volumes of lysis buffer (1 mM EDTA, 10 mM $KHCO_3$, 155mM NH_4Cl, pH 7.4). After centrifugation, the cell pellet was washed with phosphate buffered saline (Wevrick and Francke 1996) and treated as above. DNA extraction from mouse tissues was performed with the DNeasy kit (#69506, Qiagen Inc).

PCR was performed in a PTC-100/200 thermocycler (MJ Research) with reagents from Invitrogen. Reactions are in a 20 μl volume in thin walled PCR tubes. Primer sequences and conditions are in (Table 2-1).

Polymorphisms from regions of interest in human and mice strains (C57BL/6 and CAST) were identified by direct sequencing of PCR products and used to distinguish parental alleles.

RNA extraction and RT-PCR

RNA was extracted from tissues and cultured cells with Trizol reagent (Invitrogen) according to manufacturer protocols (Chomczynski and Sacchi 1987). Briefly, Trizol is an acid guanidinium thiocyanate and phenol solution that disrupts cellular membranes and proteins. Isolation of RNA is accomplished by chloroform extraction and isopropanol precipitation. Tissues required mild homogenization in a 1.5 ml eppendorf tube with a plastic disposable pestle (#199230001, ScienceWare). Total RNA was treated with DNase (Promega) and reverse transcribed (Invitrogen) into cDNA. This was used as template in

PCR reactions with gene specific-primers. Primers used in the following chapters are in (Table 2-1).

Nuclear extracts

Nuclear extracts were made by sequential lysis of cell and nuclear membranes in increasing salt (buffers NE1, 10 mM Hepes pH8, 1.5 mM MgCl2, 10 mM KCl, 1 mM DTT, and NE2, 20 mM Hepes pH8, 1.5 mM MgCl2, 25% glycerol, 420 mM NaCl, 0.2 mM EDTA, 1mM DTT, 0.5 mM PMSF) and dialyzed into NE3 (20 mM Hepes pH8, 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF). Samples were flash frozen in a dry ice methanol bath and stored at -80°C.

Electromobility shift assay (EMSA)

EMSAs were done with the help of Gareth Cory in the laboratory of Dr. Alan Underhill. 6XHIS tagged, bacterially expressed paired-domains of Pax2 and Pax8 were kindly provided by Dr. Alan Underhill. ^{32}P labeled double-stranded oligonucleotides were incubated with protein or extract of interest in EMSA buffer (2.38X stock: 50 mM Tris 7.5, 250 mM KCl, 5 mM DTT, 10 mg/ml BSA, 50% glycerol), poly[dI:dC], and allowed to bind at RT for 30 minutes. Products were separated by polyacrylamide gel electrophoresis under non-denaturing conditions (6-10%). Oligonucleotide sequences are in (Table 2-1). Gels were dried and visualized by autoradiography (Biomax MR, Kodak).

Sodium bisulfite sequencing

Sodium bisulfite sequencing was used to identify specific methylated CpG residues in DNA. During this process, unmethylated cytosine residues in the template DNA are converted to uracil, which is then converted to thymine during subsequent PCR. Bisulfite sequencing was performed as previously described (Hanel and Wevrick 2001), with minor modifications. Genomic DNA was treated with 0.3 M NaOH, 42°C/30 min, 95°C/3 min. The sample was then brought to 10X volume to 36.5% sodium bisulfite pH5 (wt/vol) and 1 mM hydroquinone and heated at 55°C overnight. The QIAEXII kit (Qiagen #20021) was used for DNA purification. PCR was performed as follows: (94°C/4 min, 58°C/2 min, and 72°C/2 min) repeated once, (94°C/30 sec, 58°C/30 sec, and 72°C/1 min) repeated 37 times, 72°C/10 min, hold at 4°C. Products were cloned into the pGEM–T vector (Promega Corp.). Ten to 30 individual clones of PCR products for each allele in each tissue were sequenced with dye terminators on an ABI sequencer. Primer sequences are in (Table 2-1).

Chromatin immunoprecipitation (ChIP)

The ChIP assay takes advantage of antibodies against specifically modified histones to enrich for DNA associated with such modifications (Figure 2-1) (Kuo and Allis 1999). While popularized by its utility in studying gene-specific histone modifications, it can also be used to study binding of non-histone chromatin proteins.

ChIP with human fibroblasts and lymphocytes

Chromatin immunoprecipitations were performed with reagents from Upstate Biotechnology (acetyl-histone H3 ChIP assay kit: #17-245, acetyl-histone H4 ChIP assay kit: #17-229, anti-dimethyl-histone H3 (Lys4): #07-030, anti-dimethyl-histone H3 (Lys9): #07-212, anti-dimethyl-histone H3 (K79): #07-366, ChIP assay kit: #17-295) and Abcam (anti-histone H3 (trimethyl K4): ab8580, anti-histone H3 (trimethyl K9): ab8898). The manufacturer's protocol was used with minor modifications. Crude lymphocyte preparations were made with Ficoll-Paque PLUS (Amersham Pharmacia Biotech) from 15 ml blood samples following the manufacturer's recommendations and expanded with phytohemagglutinin before being fixed as for other samples. LCL cultures and fibroblasts were fixed by addition of formaldehyde to a final concentration of 1% for 10 min at 37°C and cells collected. Fixed chromatin was sonicated with three 10 second pulses at one quarter maximum power with a 2 mm tip on a Fisher Scientific Sonic Dismembrator 60. Samples were pre-cleared with protein-A agarose beads prior to antibody addition. Mock control runs with no antibody were done and routinely gave no products from any primers used in this study. After reversal of the cross-links, DNA was extracted using commercially available binding columns (QIAquick PCR Purification Kit, Qiagen Inc.). The size of the resulting DNA fragments was determined by agarose gel electrophoresis. PCR was performed with one μl of template in each 20 μl reaction. *SNRPN* exon 1/IC primers were pair "A" as previously published (Saitoh and Wada 2000). PCRx (Invitrogen) was used as a PCR enhancer with a subset of primer pairs (see Supplementary Table). Quantification of ChIP experiments was done by densitometric analysis of ^{32}P end-labeled oligonucleotide probe hybridizations of slot-blotted PCR products detected on a Molecular Dynamics Typhoon and analysed with ImageQuant 5.2 quantitation software. Band intensities were corrected to background then normalized to *GAPDH* or chromosome 16 centromeric sequence (*CEN16*) bands before calculating a paternal versus maternal allele ratio. ChIP experiments were done in quadruplicate for fibroblasts, while limited availability of patient blood allowed only duplicate analysis. PCR amplifications and detection were performed multiple times for each experiment. Primer and oligonucleotide probe sequences are in (Table 2-1).

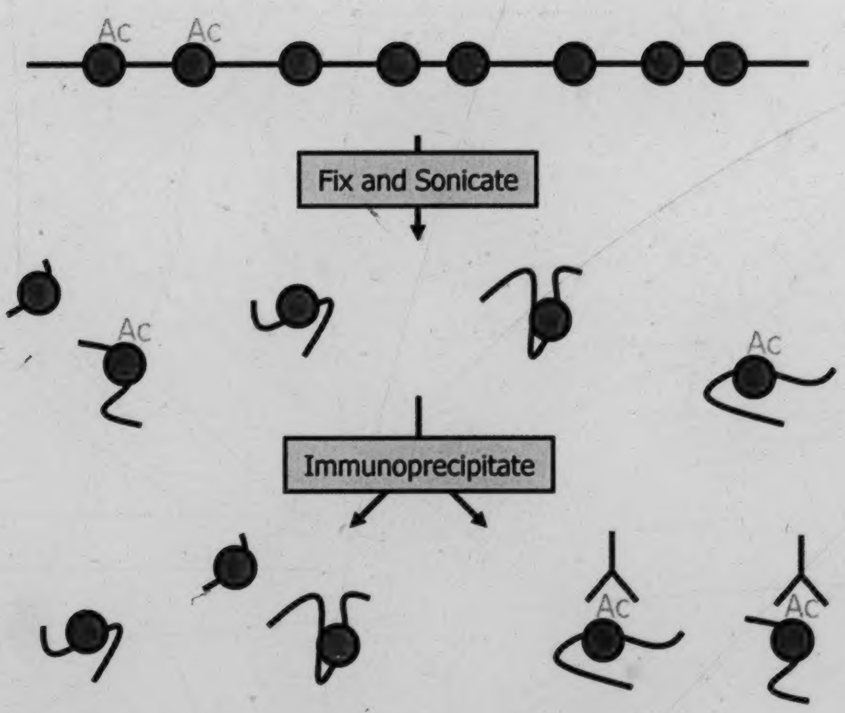

Ac
Ac
Fix and Sonicate
Ac
Ac
Immunoprecipitate
Ac
Ac

Figure 2-1. Chromatin immunoprecipitation (ChIP)
Covalent fixation between DNA and protein is caused by formaldehyde *in vivo*. Chromatin is then fragmented by sonication. Antibodies are used to enrich for nucleosomes carrying specific covalent modifications. Crosslinks caused by formaldehyde are reversed and DNA purified. Target sequences are detected by PCR amplification.

ChIP with mouse tissues

Chromatin Immunoprecipitation (ChIP) assays were performed according to manufacturer protocols (Upstate Biotech #17-295), with several modifications to the preparation of the starting material (nuclei isolation from tissues, Farnham lab, http://www.genomecenter.ucdavis.edu/farnham/farnham/protocols/tissues.html). Brain and liver were dissected from three week old F1 mice generated from a cross between a *Mus castaneus* male and a *Mus musculus* female. Tissues were chopped, homogenized with an A-pestle (loose) of a Dounce homogenizer in phosphate buffered saline (PBS) containing protease inhibitors (Mini Complete, Roche) with 1% formaldehyde, and fixed for 15 minutes at room temperature. Fixation was quenched with addition of glycine to 125 mM. Samples were further homogenized in cell lysis buffer (5 mM PIPES pH 8, 85 mM KCl, 0.5% NP40) in a Dounce homogenizer with the B-pestle (tight). Samples were strained through 3-4 layers of cheesecloth, collected by sedimentation (200 g, 5 min), and washed twice with cold PBS with protease inhibitors. Nuclei were lysed in ChIP kit lysis buffer, sonicated (twice, 10 sec each, power setting 2 on a Fisher Scientific Sonic Dismembrator 60), and specific proteins were immunoprecipitated according to a protocol from the manufacturer (Upstate). Antibodies were purchased from Upstate [anti-acetylated H3: #06-599, anti-dimethyl histone H3 (lysine 4): #07-030, anti-dimethyl histone H3 (lysine 9): #07-212, anti-dimethyl histone H3 (lysine 79): #07-366] and Abcam [anti-CTCF: ab10571, anti-trimethyl histone H3 (lysine 4): ab8580, anti-trimethyl histone H3 (lysine 9): ab8898]. Immunoprecipitated DNA was purified with the QIAEXII kit (Qiagen #20021). During the ChIP procedure, the chromatin is sonicated into fragments averaging 750 bp, so we designed PCR primers that would effectively assay each CTCF binding site cluster but that did not necessarily include the predicted binding sites themselves. PCR was performed for each region and the resulting products were slot blotted on Hybond-N membrane (Amersham Biosciences RPN303 N). Enrichment of each allele was detected by hybridization with allele-specific oligonucleotides end-labeled with ^{32}P. Quantitation of signal intensity was performed using a Typhoon 8600 (Amersham Biosciences) and ImageQuant 5.2 software from the manufacturer. Allele-specific amplification and hybridization biases were controlled by similar quantitation of input DNA, defined to be 1% of starting material. We used PCR amplification of DNA from each parental strain on each blot to control for allele specificity of probes. A Student's T-test was performed on paired data points of both alleles within the same amplification reaction with the null hypothesis that each allele is equally amplified and detected for any given template. Primer and oligonucleotide sequences are in (Table 2-1).

Modification of measurement technique

A change in the methodologies for measurement and numerical analysis were employed in Chapter 5 versus Chapter 4. In data collected for Chapter 4, a large frame of the same size was drawn around bands and the internal intensity calculated by the ImageQuant 5.2 software, with the pixels under the frame itself taken as background (local average background correction, top left example, Figure 2-2). This was sensitive to fringing effects, where non-specific labeling of edges would occur, even in wells with no sample. For data in Chapter 5, tight frames were drawn around the band only, with no background correction (top right band). Background correction was found not to be necessary because the background within a blot did not usually vary across the same blot. Shown below is an example of a blot for ChIP data in Chapter 5, with tight frames on each band.

mK4.3 16b A33.1
49

Wide border, LA correction

Narrow border, no correction

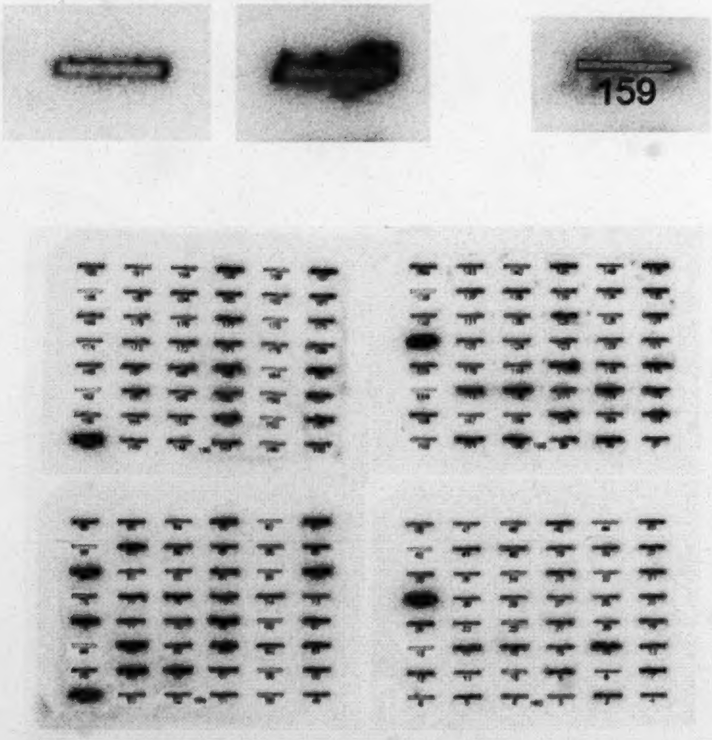


Figure 2-2. Change in measurement methods.
Example of modified measurement technique used in Chapter 5. Top left is an example of measurement used in Chapter 4 with a large frame and local average background correction. Top right is an example of measurement used in Chapter 5 with a narrow frame and no background correction. The top and middle rows show band examples described in text. Below is an example blot with narrow frames drawn.

A. CHAPTER 3 Oligonucleotides

EMSA oligonucleotides

Site	Forward Primer	F seq	Reverse primer	R seq
NDNPAX5	NDNPAX5-F	gcg cgg tcg cgc agg cgc agt gcc gcg tcc cg	NDNPAX5-R	gcg gga cgc ggc act gcg cct gcg cga ccg cg
PAX258	PAX258-F	gag ttg tga cgc act gaa gcg tga cga ctg tct	PAX258-R	gag aca gtg gtc acg ctt cag tgc gtc aca act
NDNCPBP	NDNCPBP-F	gtc ccg ccg ccg ccc cgc cct gcc cgt cgc tgc gg	NDNCPBP-R	gcc gca gcg acg ggc agg gcg ggg cgg cgg cgg ga
NDNMOK	NDNMOK-F	gca gaa caa agt aag gat ctg agc gac cct aac ttt	NDNMOK-R	gaa agt tag ggt cgc tca gat cct tac ttt gtt ctg

RTPCR primers

Amplicon	Forward primer	F seq	Reverse primer	R seq	Temp	MgCl	Cycles	Product size
PAX2	PAX2 RT-1F	cac aga cat acc ccg ttg tg	PAX2 RT-2R	tcg ttg tag gcc gtg tac tg	55	2.0	30	181
PAX5	PAX5 E5/6-F	gac gaa ggt att cag gag tc	PAX5 E10-R	cca agg gtc agt gac ggt c	55	2.0	30	585
PAX8	PAX8 RT-1F	gca acc att caa cct ccc ta	PAX8 RT-2R	ctg ctg ctg ctc tgt gag tc	55	2.0	30	240
CPBP	COPEB-1F	cac gag acc ggc tac ttc tc	COPEB-2R	ctg acc aaa act tcg cca at	55	2.0	30	359
EGR-1	EGRL-MMF	cag cag tcc cat tta ctc ag	EGRL-MMR	gac tgg tag ctg gta ttg	55	2.0	30	344
MAZR	ZNF278-1F	gcg ccg ata taa tgc tct tt	ZNF278-2R	tac cac ata gac cgc atg ga	55	2.0	30	390
hsMOK2	MOK2-1F	tcc ctg agc tgc aaa ctt ct	MOK2-2R	att gta ctg gcc atg cct tc	55	2.0	30	352

B. CHAPTER 4 Oligonucleotides

ChIP Primers

Amplicon	Forward primer	F seq	Reverse primer	R seq	Temp	MgCl	Cycles	Product size	PCRx
A	NEC122F	GGC CTA TTG CTA TGC CTG TC	NEC123R	CAC AGA GGC TGT CTC CCT TC	60	2.5	33-35	200	
B	NEC80F	TCC TCT CAC TGG TTC GCA TA	NEC81R	TGG GCT GAG AAG ATC TAG GG	55	2.0	33-35	219	
C	NEC87F	CCT GCC CTA GAT CTT CTC AGC	NEC34R	GGG GCC TCG GCT GCA AAG TTA GG	60	2.0	33-35	352	2X
D	NEC21F	GCG CAG ACA TGT CAG AAC AA	NEC69R	TTG ACC AGC ACG TAC CAC AT	50	3.0	33-35	343	3.5X
E	NEC16F	ACG AGC TCA TGT GGT ACG TG	NEC17R	GAA GGT GGA GTG CTT CTT CC	50	2.5	33-35	376	2X
F	NEC20F	GCC CGA ATA CGA GTT CTT TT	NEC6R	CAC ACA TCA TCA GTC CCA TA	55	2.5	33-35	540	
G	NEC120F	TGT GAG CAC TTG GCA CAC TT	NEC121R	GCG ATT TTT CCC ACC CTA TT	50	2.0	33-35	211	
H	NEC137F	GGC AGA AAA ACA ATG GAA GC	NEC138R	TTG TTT CTT TGT ACT ATT TTT CCT TTC	60	2.5	33-35	117	
SNRPN A	SNRPN AF	GAT GCT CAG GCG GGG ATG TGT GCG	SNRPN AR	GCT CCC CAG GCT GTC TCT TGA GAG	60	2.5	33-35	172	
GAPDH	GAPDH F	GCA TCA CCC GGA GAA ATC GG	GAPDH R	GTC ACG TGT CGC AGA GGA GC	60	2.5	33-35	268	
CEN16	CEN16 F	GTC TCT TTC TTG TTT TTA AGC TGG G	CEN16 R	TGA GCT CAT TGA GAC ATT TGG	55	2.5	33-35	207	

ChIP Oligonucleotide Probes

A	NEC176	cac cac caa aag ccc ttt ta
B	NEC177	gag aag ggg cca gtt taa gg
C	NEC178	gca aag tta ggg tcg ctc ag
D	NEC68	GAG CGA CCC TAA CTT TGC AG
E	NEC14	ATG CTC CTG CAC CAC TTC TT
F	NEC2	ATT TGC ATC TTG GTG ATT T
G	NEC8	GGT GGG GTT GTA TAT GTG TT
H	NEC179	ttg aaa caa gtt ttt gct tcc a
SNRPN A	SNRPN A O	tga cgc atc tgt ctg agg ag
GAPDH	end-labeled PCR product	
CEN16	end-labeled PCR product	

Table 2-1. Tables of oligonucleotides.
Oligos used in various experiments in (A) Chapter 3, (B) Chapter 4, (C) Chapter5.

C. CHAPTER 5 Oligonucleotides

ChIP Primers

		size	MgCl	temp	cycle
Frat3	cca ttc agt ggg tgt cag aa	222	2.5	60	33-35X
	aga atg ggc atc tga gac aa	222	2.5	60	33-35X
Mkrn3	gac agt gtc cct gcc aaa g	205	2.5	60	33-35X
	ggc aaa gtc cag gct tct ac	205	2.5	60	33-35X
Ndn	cca tca tct agt tct gtg cca	284	1.5	60	33-35X
	ctt cgg atc aga gca gga c	284	1.5	60	33-35X
iCT-A	aag tgg ttg gcc ttg tct gt	142	2.5	60	33-35X
	ttc aga atg aca gtt cac att gc	142	2.5	60	33-35X
iCT-B	agg aac act tgt ggc ttg aga	217	2.5	60	33-35X
	gaa aca cac tgc agc agc tc	217	2.5	60	33-35X
Snurf/Snrpn/IC	caa cag agc tcc tgc atc ct	226	2.5	60	33-35X
	gcc tct gga ctc ctg gaa g	226	2.5	60	33-35X
Dadl ex1'	ggg cag cag tac tcc acc aa	159	2.5	60	33-35X
	cgt agg atg cag gga ttt tct tta	159	2.5	60	33-35X
Dadl ex6	ccc aca gat tga aca cag gaa at	179	2.5	60	33-35X
	gag gga tgg atg tct gct gtt t	179	2.5	60	33-35X

RT-PCR Primers (Frat3 and Mkrn3 requires PCRx)

		size	MgCl	temp	cycle	PCRx
Frat3	gac gtg gac cgg ctc atc	509	2.5	55	30-35X	2X
	agg ttt ccc gaa aga agg ag	509	2.5	55	30-35X	2X
Mkrn3	caa gcc ttg cag cag gtg	298	2.5	60	30-35X	1X
	att tct cca tgg ggg tat gc	298	2.5	60	30-35X	1X
Ndn	gta tcc caa atc cac agt gc	356	2	55	30-35X	
	ctt cct gtg cca gtt gaa gt	356	2	55	30-35X	
iCT-A	cag cag agg act tcc tgg tc	150	1.5	60	30-35X	
	tgg ggt gtt aat tcc acg tt	150	1.5	60	30-35X	
Snurf/Snrpn/IC	CCC GAG TAT TAA GGA TCT TG	142	1.5	55	30-35X	
	TGA AGA TTC TCC CAT CTT GC	142	1.5	55	30-35X	
Gapdh	GCC ATC AAC AAC CCC TTC AT	315	2	60	30-35X	
	TTC ACA CCC ATC ACA AAC AT	315	2	60	30-35X	

Allele-specific oligonucleotide probes

		allele
Frat3	tct taa agt cag att aca g	Cast
	tct taa agt aag att aca g	C57
Mkrn3	cct tcg acg ctt gta	Cast
	cct tca atg ctt gta	C57
Ndn	tgc ttc gct cct ttc	C57
	tgc ttc ggt cct ttc	Cast
iCT-A	cgg cct cca tct ag	Cast
	cgg ctt cca tct ag	C57
iCT-B	ccc tgc cta gcc ct	C57
	ccc tgt cta gcc ct	Cast
Snurf/Snrpn/IC	aac ttc tac cca cac cc	Cast
	aac ttc tat cca cac cc	C57
Dadl ex1'	ctc ggc gtg ttt att c	
Dadl ex6	cac aca cac aca gag gag	

Bisulfite Sequencing Primers

Frat3 required semi-nested PCR, first round with F-flank/R and second round with F-nest/R

			size
Frat3	gag ttt ttt tgg tgg taa tga tta ga	F-flank	746
	ttt taa tta aga atg aga agt tta ggt tt	F-nest	676
	aaa tac tta ctt tac cca tcc cc	R	
Mkrn3	ggg taa ttg aat ttg ttt ttg gat a		238
	taa aaa tat aca cac cta tcc cca c		238
iCT-A	aaa aaa act ccc aca aca caa taa c		312
	ttt tta ggt tta gtt aaa ttt ttt aga tta		312

Chapter 3 ♦ Analysis of the *NDN* Promoter

Parts of this chapter have appeared in:

Hanel ML, **Lau JC**, Paradis I, Drouin R, Wevrick R (2005) Chromatin modification of the human imprinted *NDN* (*necdin*) gene detected by *in vivo* footprinting. J Cell Biochem 94(5):1046-57

Introduction

Gene regulation involves complex networks of *cis*-acting genetic elements and *trans*-acting protein factors. The most tangible of these *cis*-acting elements is the promoter of a gene. Binding and activation of RNA polymerase and associated transcriptional machinery is the ultimate end result of a concert of regulatory events that includes the availability of tissue-specific transcription factors as well as accessibility of the locus in terms of its chromatin state. For imprinted genes that are also transcribed in specific spatial-temporal patterns, such as *NDN*, *MAGEL2*, and *MKRN3*, an extra layer of regulation must intersect at the promoter to control correct allele-specific expression. Detailed examination of the promoter of these genes may give clues as to chromatin accessibility as well as the *trans*-acting factors involved.

NDN is expressed at varying levels in many human tissues but is highest in brain, heart, muscle and fibroblasts. However, its expression is not ubiquitous and *NDN* is not expressed at appreciable levels in liver, kidney and blood leukocytes (Jay et al. 1997; MacDonald and Wevrick 1997). Mouse *Ndn* is also tissue-specific, and is expressed in a pattern that partially overlaps the human *NDN* expression profile (Uetsuki et al. 1996). In particular, it is more neurally restricted. Since *NDN* was localized to the PWS critical region, it became an excellent candidate gene for phenotypes in the disease because it had been implicated in terminal differentiation of neurons (Uetsuki et al. 1996; MacDonald and Wevrick 1997), respiration (Ren et al. 2003), and axon outgrowth (Lee et al. 2005). While the expression profiles of human and mouse *NDN/Ndn* are not identical, the similarities suggest overlap in aspects of tissue-specific regulation. Both orthologues are imprinted in all expressed tissues examined, and therefore may share mechanisms of allele-specific expression.

Previous analysis of the human and mouse *NDN/Ndn* promoters involved reporter construct transfections into cultured cells (Uetsuki et al. 1996; Nakada et al. 1998). This group defined the minimal promoter in both species and found that the mouse promoter conferred expression of a reporter gene in post-mitotic neurons during *in vitro* differentiation of P19 cells. It was also found that the human promoter was sensitive to DNA methylation, in that methylation of seven *HhaI* sites greatly reduced the ability of this promoter to activate the reporter. In a comparison of the DNA sequences of the two promoters, a site resembling the binding site for the transcription factor SP1 was found in both sequences in corresponding locations (Nakada et al. 1998). In this chapter, I further analyzed the DNA sequence of the promoters and characterized other potentially important sites where *trans*-acting factors may bind to mediate tissue or allele-specific transcription.

Footprinted sequences in the NDN promoter

In parallel to this work, Dr. Meredith Hanel of our laboratory had undertaken *in vivo* footprinting studies of the *NDN* promoter (Hanel et al. 2005). Briefly, deletion patient cell lines from PWS and AS patients were used in DNase I, DMS and UVC footprinting assays. The use of deletion patient cells allowed unequivocal differentiation of alleles, with the maternal allele remaining in PWS, and paternal allele remaining in AS. Fibroblasts and lymphocytes were used to represent *NDN* expressing and non-expressing tissues, respectively. These studies enabled a fine scale ascertainment of chromatin access on either allele in expressing and non-expressing contexts. I used the data generated by Dr. Hanel to examine promoter occupancy, in comparison with bioinformatics analysis of the promoter, to identify candidate factors regulating allele-specific and tissue-specific transcription of *NDN*.

One of the footprints Dr. Hanel found was a striking paternal allele-specific DNase I footprint on both strands, located in the promoter between –136 and –166 bp upstream of the start codon (Hanel et al. 2005). This was only found in fibroblasts, and therefore may correspond to transcription-related factors used by the expressing allele. The occupancy of this site was supported by the presence of UVC and DMS footprints as well. This indicates that a protein or protein complex is bound to the active allele, covering both strands over a region of about 30 bp with high GC content. This footprint will be referred to as the transcription-related footprint.

Conservation of sequence features between these promoters despite the 75 million years between human and mouse (Waterston et al. 2002) suggests a functional role and will lead to a better understanding of the regulation of *NDN/Ndn* (Pennacchio and Rubin 2001). To this end, this chapter will describe comparative sequence analysis of the *NDN/Ndn* promoter of human, mouse, and rat. These phylogenetic footprints were correlated with *in vivo* footprints performed by Dr. Hanel. Overlaps between phylogenetic and *in vivo* footprints are of special interest, and were investigated in terms of sequence motifs previously characterized to be binding sites of transcription factors. Potential binding of a candidate factor was studied using *in vitro* binding assays. The data suggest that the differential footprinted patterns observed between expressed and non-expressed alleles may be due in part to accessibility of the chromatin at the promoter region and not solely to the availability of *trans*-acting factors.

Results

Phylogenetic analysis of the NDN/Ndn promoter

To study the sequence conservation between human, mouse and rat *NDN/Ndn*, which may be indicative of conserved functional elements, the upstream regions from the three species were compared using ClustalW as implemented by the European Bioinformatics Institute (Figure 3-1) (Thompson et al. 1994). By visual inspection, it is clear that the degree of similarity between mouse and rat sequences is very high, as expected for the closely related species. To further quantitate the sequence similarity between the three sequences, a percentage identity plot was produced from the ClustalW data, using a 50 bp window (Figure 3-2 A). As can be seen from the plot of the alignment between all three sequences, the similarity drops dramatically a short distance upstream of the transcription start. Since the mouse and rat sequences retain more extensive sequence similarity, the change in similarity among all three sequences is attributable to differences between human and rodent sequences. Closer inspection of the sequence alignment (Figure 3-1) shows a well defined point at which the level of similarity changes at position -133. In fact, overall sequence similarity 5' to this is 40.5%, while it is 69.1% 3' of this position, perhaps suggesting an evolutionary rearrangement event between human and rodents. It is notable here that most of the previously defined minimal promoters in human and mouse are in the less conserved region (Figure 3-2 A).

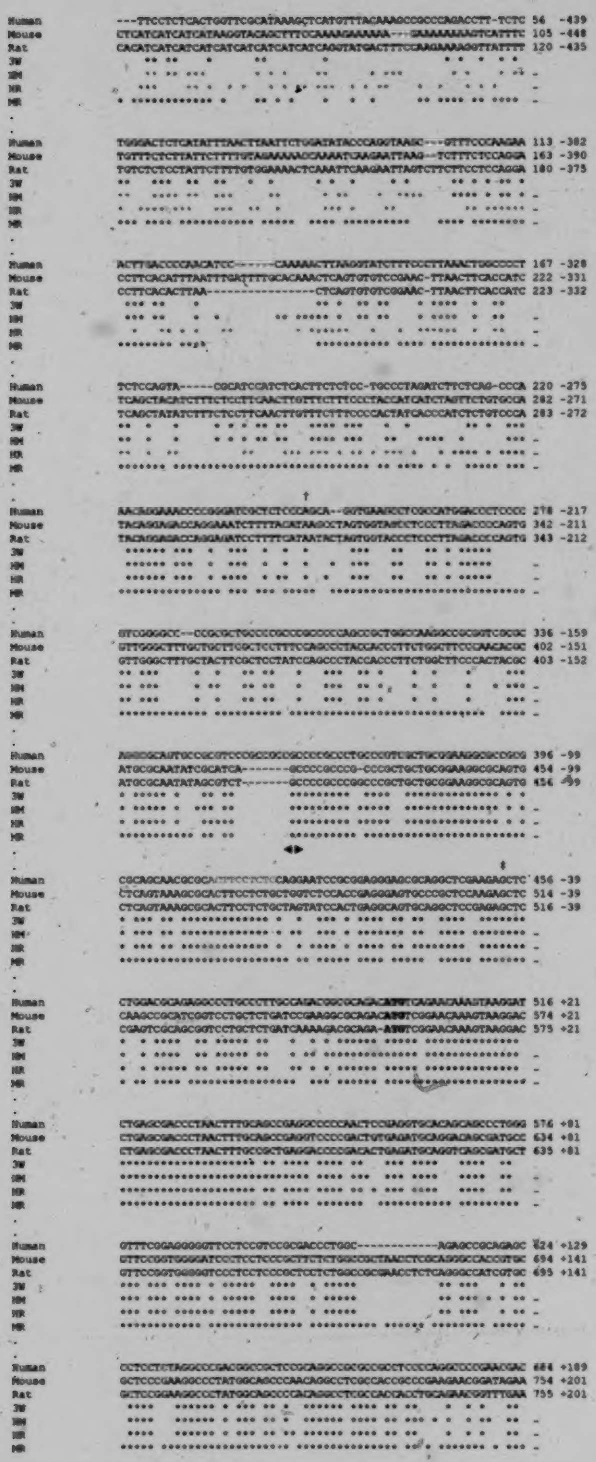
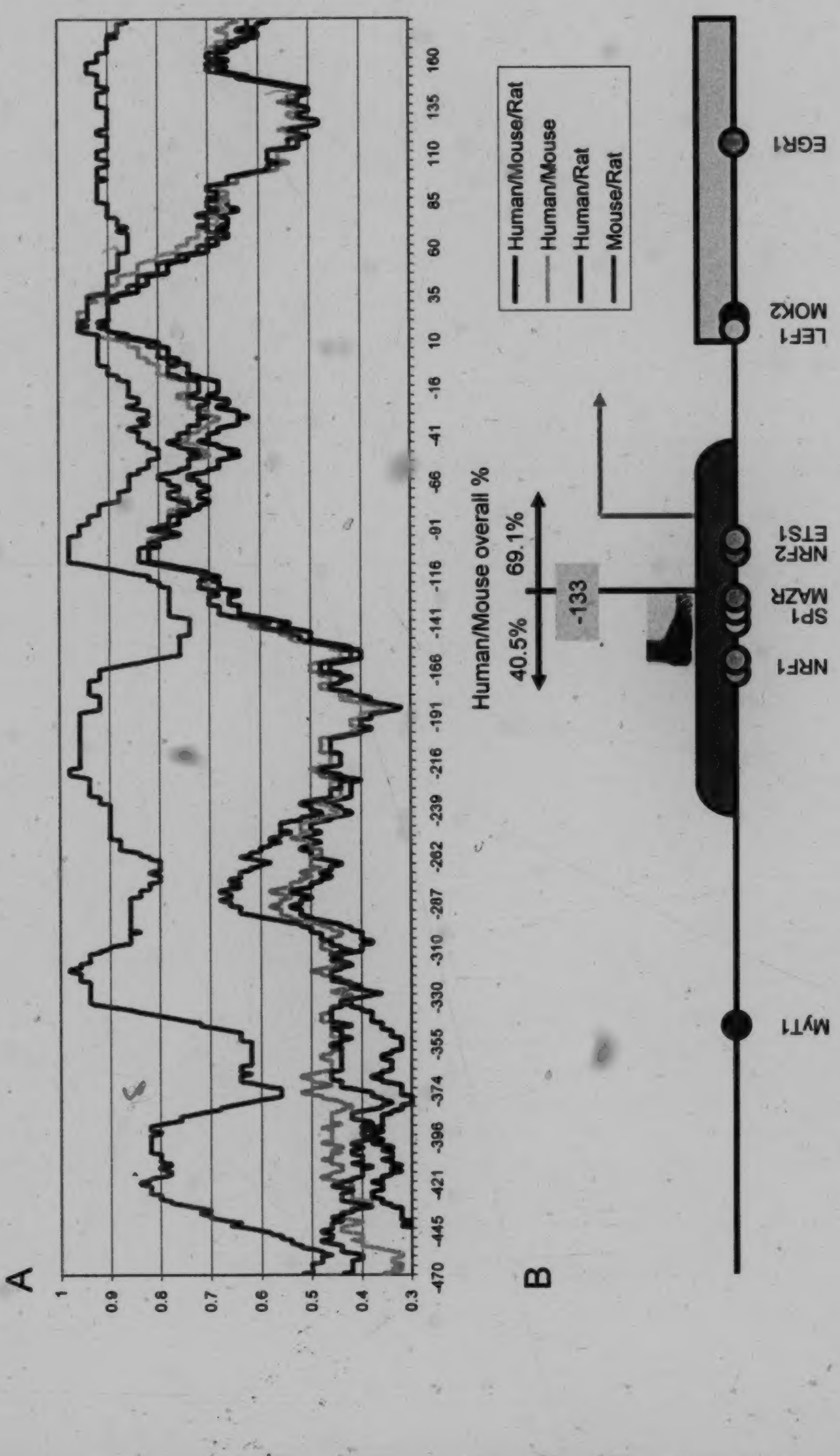


Figure 3-1. Sequence alignment of *NDN/Ndn*. ClustalW alignment of human, mouse and rat *NDN/Ndn*. Multiple alignment with all three sequences is shown. Asterisks indicate sequence identity across all three (3W), human versus mouse (HM), human versus rat (HR), and mouse versus rat (MR). Red asterisks indicate approximate location of identity in another frame with respect to the multiple alignments. The start codon is indicated in blue, the transcription start indicated in green, the minimal promoter between single and double daggers, and the two opposing arrowheads indicates division between weak and strong sequence conservation between human and mouse.

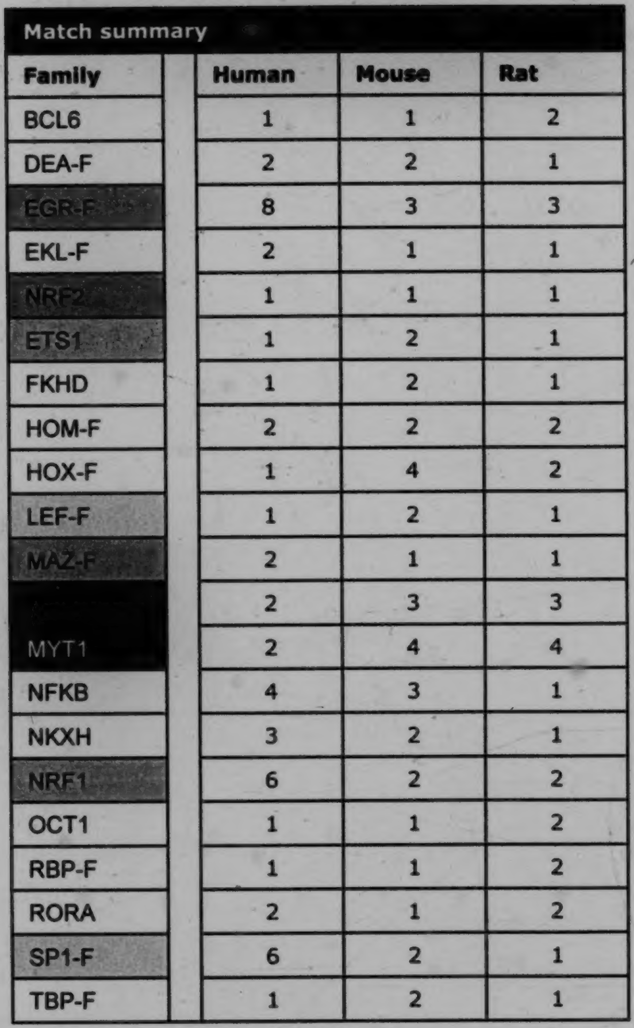
Figure 3-2. Percentage identity plot and predicted TF sites.

(A) Percentage identity plots for the upstream region of *NDN/Ndn* in human, mouse and rat. The Y-axis indicates percentage identity across 50 bp windows. X-axis indicates the position, numbered with start codon as +1. (B) Schematic showing features upstream of *NDN/Ndn*. Dark green rounded bar indicates minimal promoter. Green arrow indicates transcription start. Light green rectangular bar indicates ORF. Vertical black line indicates position -133, a division between weak and strong sequence conservation between human and mouse. Foot indicates the transcription-related *in vivo* footprint in human *NDN*. Small coloured circles indicate position of consensus binding sites for transcription factors found across all three species and at the same position.


A
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
-470
-445
-421
-396
-374
-355
-330
-310
-287
-262
-239
-216
-191
-166
-141
-116
-91
-66
-41
-16
10
35
60
85
110
135
160
Human/Mouse/Rat
Human/Mouse
Human/Rat
Mouse/Rat
B
Human/Mouse overall %
40.5%
69.1%
-133
MyT1
NRF1
SP1
MAZR
NRF2
ETS1
LEF1
MOK2
EGR1

The sequence of the transcription-related footprint characterized by Dr. Hanel was analyzed with MatInspector to find putative transcription factor binding consensus sites (Quandt et al. 1995). Within this region are consensus sites for MAZR (MYC-associated zinc finger protein related transcription factor, ZNF278) (Kobayashi et al. 2000) and a pair each of SP1 and NRF1 (Scarpulla 2002) sites that are conserved with mouse and rat with respect to position. Specific to the human upstream sequence are also additional NRF1, AP2, PAX5, SP2, EGR1, and CPBP sites (Figure 3-3) (Koritschoner et al. 1997; Busslinger 2004; Simon et al. 2004).

To further characterize conservation of sequence motifs that may be indicative of transcription factor binding sites, sequences upstream of *NDN/Ndn* were analyzed. Numerous sequences with similarity to transcription factor consensus sites were found in the human, mouse and rat sequences (112, 103, and 88 respectively). Putative factor binding sites significant to regulation of *NDN/Ndn* are expected to be conserved across all three species (Table 3-1). Furthermore, conservation in position would also be suggestive of conservation of an important factor binding site (Figure 3-2 B). Consensus sites for nine transcription factors were found in the promoter sequences of all three species, and in a similar position with respect to the start codon. Of the nine putative factors, there were two closely spaced consensus sites each for NRF1 and SP1 (Figure 3-3).



Match summary

Family	Human	Mouse	Rat
BCL6	1	1	2
DEA-F	2	2	1
EGR-F	8	3	3
EKL-F	2	1	1
NRF2	1	1	1
ETS1	1	2	1
FKHD	1	2	1
HOM-F	2	2	2
HOX-F	1	4	2
LEF-F	1	2	1
MAZ-F	2	1	1
[illegible]	2	3	3
MYT1	2	4	4
NFKB	4	3	1
NKXH	3	2	1
NRF1	6	2	2
OCT1	1	1	2
RBP-F	1	1	2
RORA	2	1	2
SP1-F	6	2	1
TBP-F	1	2	1

Table 3-1. Table of predicted TFs.
Summary of putative transcription factor sites shared between human, mouse and rat. Numbers indicate multiples of those sites found in a particular species. Colored background corresponds to factors in Figure 3-2 B.

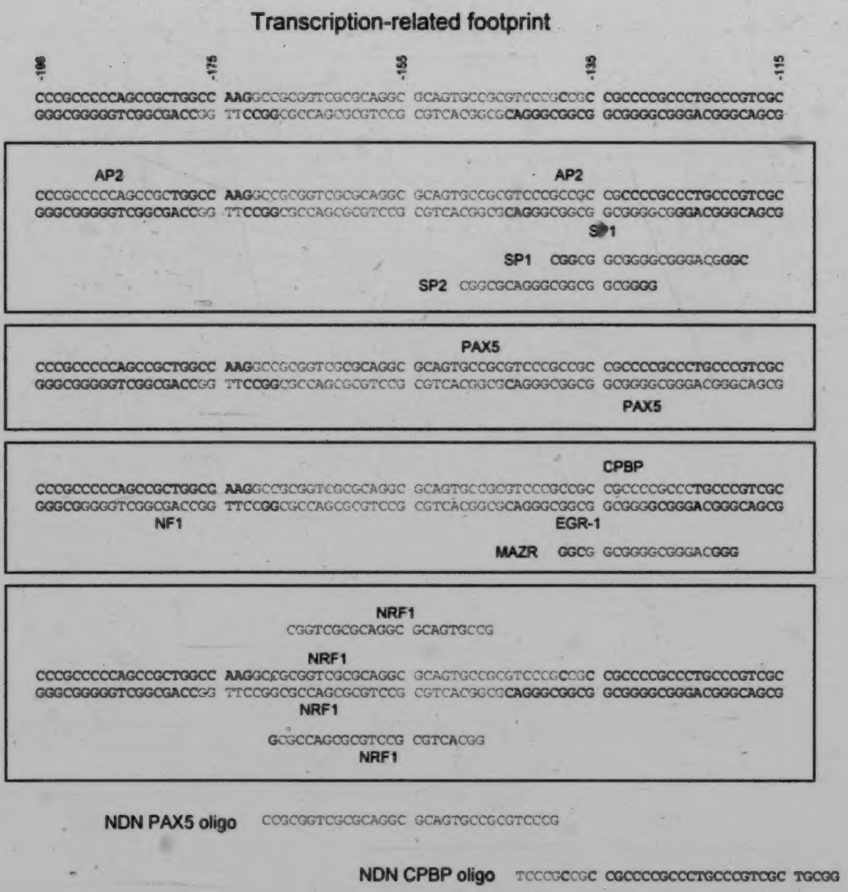

Transcription-related footprint
-196
-175
-155
-135
-115
CCCGCCCCCAGCCGCTGGCC AAGGCCGCGGTCGCGCAGGC GCAGTGCCGCGTCCCGCCGC CGCCCCGCCCTGCCCGTCGC
GGGCGGGGGTCGGCGACCGG TTCCGGCGCCAGCGCGTCCG CGTCACGGCGCAGGGCGGCG GCGGGGCGGGACGGGCAGCG
AP2
AP2
CCCGCCCCCAGCCGCTGGCC AAGGCCGCGGTCGCGCAGGC GCAGTGCCGCGTCCCGCCGC CGCCCCGCCCTGCCCGTCGC
GGGCGGGGGTCGGCGACCGG TTCCGGCGCCAGCGCGTCCG CGTCACGGCGCAGGGCGGCG GCGGGGCGGGACGGGCAGCG
SP1
SP1 CGGCG GCGGGGCGGGACGGGC
SP2 CGGCGCAGGGCGGCG GCGGGG
PAX5
CCCGCCCCCAGCCGCTGGCC AAGGCCGCGGTCGCGCAGGC GCAGTGCCGCGTCCCGCCGC CGCCCCGCCCTGCCCGTCGC
GGGCGGGGGTCGGCGACCGG TTCCGGCGCCAGCGCGTCCG CGTCACGGCGCAGGGCGGCG GCGGGGCGGGACGGGCAGCG
PAX5
CPBP
CCCGCCCCCAGCCGCTGGCC AAGGCCGCGGTCGCGCAGGC GCAGTGCCGCGTCCCGCCGC CGCCCCGCCCTGCCCGTCGC
GGGCGGGGGTCGGCGACCGG TTCCGGCGCCAGCGCGTCCG CGTCACGGCGCAGGGCGGCG GCGGGGCGGGACGGGCAGCG
NF1
EGR-1
MAZR GGCG GCGGGGCGGGACGGG
NRF1
CGGTCGCGCAGGC GCAGTGCCG
NRF1
CCCGCCCCCAGCCGCTGGCC AAGGCCGCGGTCGCGCAGGC GCAGTGCCGCGTCCCGCCGC CGCCCCGCCCTGCCCGTCGC
GGGCGGGGGTCGGCGACCGG TTCCGGCGCCAGCGCGTCCG CGTCACGGCGCAGGGCGGCG GCGGGGCGGGACGGGCAGCG
NRF1
GCGCCAGCGCGTCCG CGTCACGG
NRF1
NDN PAX5 oligo CCGCGGTCGCGCAGGC GCAGTGCCGCGTCCCG
NDN CPBP oligo TCCCGCCGC CGCCCCGCCCTGCCCGTCGC TGCGG

Figure 3-3. Predicted factor binding sites in transcription-related footprint.

Shown at top is the DNA duplex of positions -196 to -115 upstream of human *NDN* start codon. In blue, is the transcription-related footprint found on the expressing allele in the expressing tissue. Below are the same sequences, with putative transcription factor binding sites shown in red.

Verification of bioinformatically predicted PAX5 site by EMSAs

While binding sequences for NRF1 are coincident with the transcription-related footprint, making it a possible candidate for a factor occupying the footprint, the presence of these sites was not known until the latest update of the MatInspector database (version 7.4, Jan, 2005). The experiments described here are based on data from a previous database revision (6.1, Jan, 2003). The next best candidate for a factor responsible for the footprint is PAX5. Other than the two consensus sites found near the footprint, the human sequence has a third consensus site at position -216. To test if PAX5 can indeed bind the predicted sequence upstream of *NDN*, EMSAs were performed with the aid of Gareth Cory of the laboratory of Dr. Alan Underhill. Radiolabeled ds-oligonucleotides were incubated with proteins of interest to form protein-DNA complexes, which were detected after size separation on polyacrylamide gel electrophoresis. PAX5 binds DNA through its paired-domain, and has highly similar binding specificity to the other members of its subfamily, PAX2 and PAX8; therefore paired-domains from these proteins were used to test binding to upstream regions of *NDN* (Figure 3-4). These domains were produced in bacteria from constructs that also encoded a 6XHIS tag for purification purposes. The binding of these purified Pax2 and Pax8 paired-domains to double-stranded DNA representing the promoter region of *NDN* found to be occupied by the transcription-related footprint (positions -170 to -139, NDNPAX5) was tested. Both the Pax2 and Pax8 paired-domains were able to bind this 30 bp double stranded oligonucleotide NDNPAX5, although not as well as to an idealized consensus sequence of the same length bound by the PAX2/5/8 subfamily of paired-domains (PAX258, (Czerny and Busslinger 1995)) (Figure 3-4). The Pax2 paired domain also had a qualitatively higher affinity to the NDNPAX5 sequence than Pax8 paired domain.

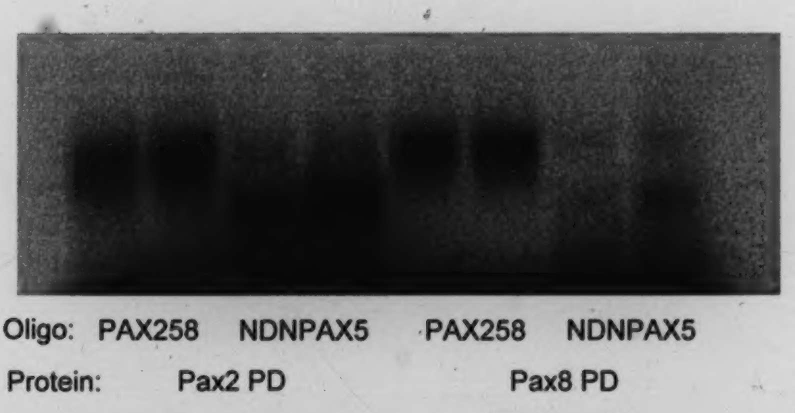

Oligo: PAX258 NDNPAX5 PAX258 NDNPAX5
Protein: Pax2 PD Pax8 PD

Figure 3-4. EMSA of footprinted site to related paired domains. EMSA of labeled oligos made to transcription-related footprinted region in *NDN* promoter binding to Pax2 and Pax8 paired domains, which are highly related to the Pax5 paired domain. Oligo PAX258 is of an idealized binding sequence, and NDNPAX5 is of a 30b sequence occupied by the transcription-related footprint.

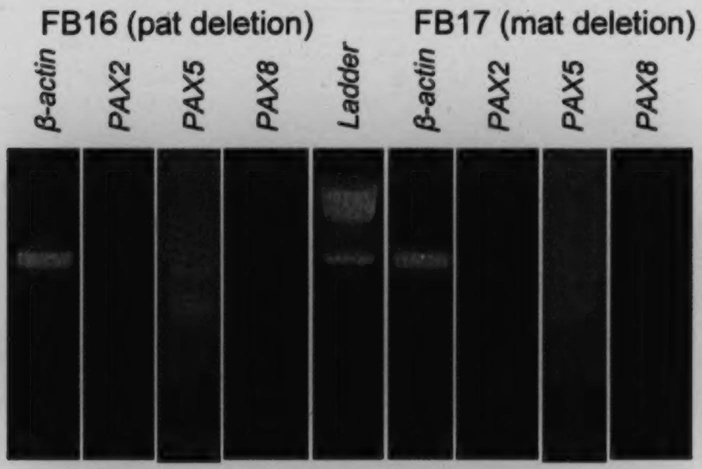

FB16 (pat deletion)
FB17 (mat deletion)
β-actin
PAX2
PAX5
PAX8
Ladder
β-actin
PAX2
PAX5
PAX8

Figure 3-5. RT-PCR of *PAX* genes in fibroblast lines. RT-PCR of *PAX2*, *PAX5* and *PAX8* in PWS and AS patient fibroblasts (FB16 and FB17, respectively).

In humans, these proteins are normally expressed postnatally only in retina and kidney (*PAX2*) (Eccles et al. 2002; Pichaud and Desplan 2002), B-cell lineages (*PAX5*) (Borson et al. 2002) and thyroid (*PAX8*) (Christophe 2004). To further investigate the possibility that one of the *PAX2/5/8* subfamily of transcription factors may be expressed in cells where the footprint was found, we performed reverse transcription-PCR for these three PAX genes in the fibroblast cell lines used in the *in vivo* footprinting and found no evidence for expression of *PAX2*, *PAX5* or *PAX* 8 (Figure 3-5). This transcription-related footprint may thus identify a binding site for another human transcription factor such as NRF1, or an unidentified one related in binding specificity to proteins of the PAX family.

Allelic differences in footprints are due to chromatin states

It was necessary to determine if allelic footprints were due to differences between alleles in terms of chromatin accessibility or due to differences in complement of proteins capable of binding in the individual cell lines. Therefore, I tested if similar proteins are in each cell line that has the *in vitro* potential to bind the promoter sequences. Nuclear extracts were prepared from PWS and AS deletion fibroblasts and lymphoblasts and binding to *NDN* upstream sequences was assayed. The double-stranded oligonucleotide NDNPAX5 corresponds to the transcription-related footprint found on the paternal expressed allele of the fibroblast, while the ds-oligonucleotide NDNCPBP corresponds to a region containing the predicted CPBP, EGR-1 and MAZR sites (position -148 to -110) (Figure 3-2). Ds-oligonucleotide NDNMOK corresponds to a region (+5 to +40) coinciding with a maternal allele-specific DNase I footprint found in both fibroblasts and lymphoblasts, and is coincident with a predicted MOK-2 consensus site. Similar to previous EMSA experiments, radiolabeled ds-oligonucleotides were incubated with nuclear extracts and size separated to give patterns indicative of protein-DNA complexes. As expected, we found no evidence that there are differences in the complement of factors able to bind the oligonucleotides between PWS and AS cell lines (Figure 3-6). This indicates there are factors available in the PWS fibroblasts that can bind to the footprinted sequence, but do not, presumably due to inaccessibility of the chromatin on the maternal allele. DNA methylation may also play a role here to exclude binding of factors, as it has been shown that the *NDN* promoter is more methylated on the maternal allele in this region (Lau et al. 2004).

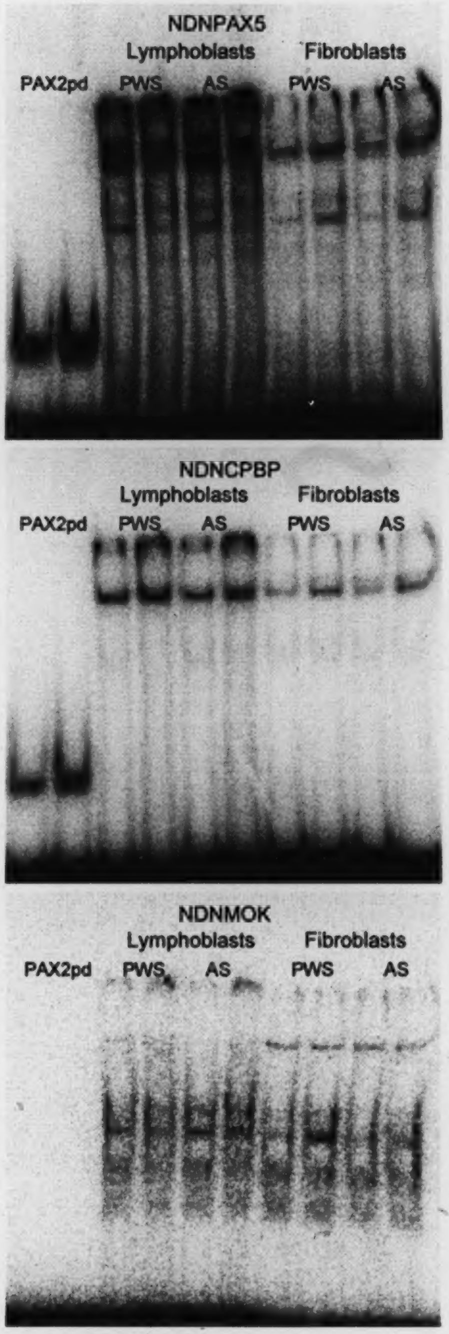

NDNPAX5
Lymphoblasts
Fibroblasts
PAX2pd
PWS
AS
PWS
AS
NDNCPBP
Lymphoblasts
Fibroblasts
PAX2pd
PWS
AS
PWS
AS
NDNMOK
Lymphoblasts
Fibroblasts
PAX2pd
PWS
AS
PWS
AS

Figure 3-6. EMSAs with patient cell line nuclear extracts.
EMSAs with nuclear extracts from PWS and AS fibroblasts and lymphoblasts. Pax2 paired-domain (PAX2pd) included for comparison. NDNPAX5 corresponds to the transcription-related footprint in AS fibroblasts, NDNCPBP corresponds to predicted factor sites downstream of the footprint, and NDNMOK corresponds to MOK-2 putative site downstream of the promoter. For each pair of EMSA reactions, the left and right lanes indicate addition of 5μl and 10μl of extract, respectively.

To investigate the presence of specific candidate factors that may bind the *NDN* promoter in both our fibroblast lines, expression of *CPBP*, *EGR-1*, *MAZR* and *MOK-2* were tested by RT-PCR (Figure 3-7). We found no evidence that there are differences in the expression of candidate factors between PWS and AS cell lines. This is consistent with the nuclear extract EMSA data above, which does not show any differences in binding patterns of oligonucleotides between PWS and AS cells. In contrast to the *PAX2/5/8* family, *CPBP* and *EGR-1* are strongly expressed in both fibroblast lines. *MOK-2* is expressed at limited levels, while *MAZR* expression is not detectable.

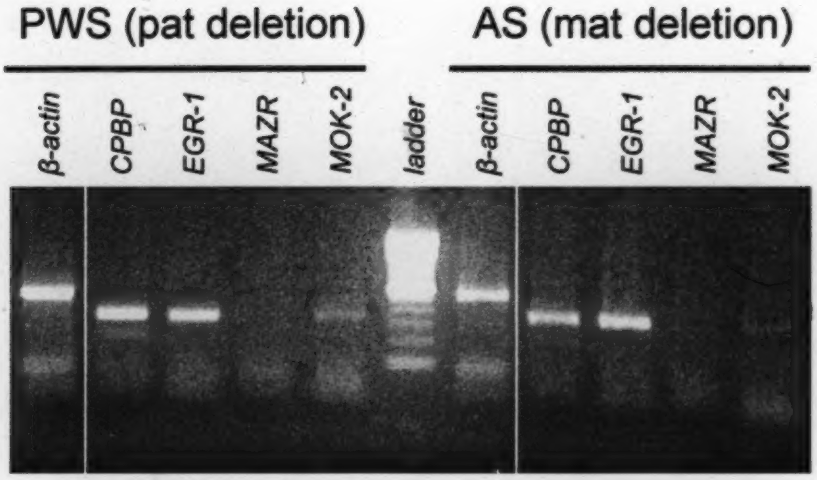

PWS (pat deletion)
AS (mat deletion)
β-actin
CPBP
EGR-1
MAZR
MOK-2
ladder
β-actin
CPBP
EGR-1
MAZR
MOK-2

Figure 3-7. RT-PCR of candidate factors binding to the *NDN* promoter.
RT-PCR of candidate binding factors in PWS and AS patient fibroblasts.

Discussion

The regulation of *NDN* is influenced by multiple factors including chromatin state and *trans*-acting factors that direct its correct allelic and developmental expression. The *in silico* approach to predict possible transcription factor binding sites and the complementary *in vivo* data from the work of Dr. Meredith Hanel has given insight into a cross-section of these layers. The bioinformatics analysis has shown several putative conserved factor binding sites in regions of high cross-species sequence conservation, and surprisingly, also in regions of little conservation between human and rodent (Figure 3-2). Furthermore, investigation of the *in vivo* occupancy of the human *NDN* promoter by footprinting has aided identification of potentially important regulatory elements without cross-species conservation. Similar approaches have led to functional characterization of factors regulating other genes (Smith et al. 2004b).

The bioinformatic prediction of transcription factor binding sites suggested several motifs were found across human, mouse and rat, and in a similar position relative to the open reading frame of *NDN/Ndn*. Each transcription factor may have a role in regulating the tissue or allele-specificity of transcription. MyT1 is a transcription factor that is thought to interact with neurogenin 1 and specifies vertebrate early neuronal precursor fates by affecting the *Notch* signaling pathway (Quan et al. 2004). This is consistent with a role for *NDN* in brain function (Ren et al. 2003; Lee et al. 2005). It is also notable that there are multiple possible MyT1 sites found in human, mouse and rat, one of which is conserved in position to *NDN/Ndn* in a region of little overall sequence conservation, which suggests it may be a functional component of the regulatory elements controlling *NDN/Ndn*.

NRF-1 and SP1 sites are also found upstream of *NDN/Ndn* in all three species. NRF (nuclear respiratory factor) proteins function to regulate nuclear genes important in mitochondrial biogenesis and function (Scarpulla 2002). While *NDN* has no previously characterized role in mitochondrial function or energy metabolism, NRF-1 has been shown to be involved in regulation of *FMR1*, a gene that has also not been shown to be immediately involved in mitochondrial function (Smith et al. 2004a). In fact, at the *FMR1* promoter, NRF-1 synergistically interacts with SP1, whose binding sites are also predicted to be in close proximity in the promoter of *NDN*. It will be of interest to investigate whether the same interaction and activation between NRF-1 and SP1 also occurs to regulate *NDN*.

Little is known about the function of MAZR, although it interacts with Bach2, a B-cell and neuron-specific transcriptional repressor (Kobayashi et al. 2000). A neuron-specific function would be of interest to *NDN* biology, but a role in B-cells has yet been suggested for *NDN*. Like MAZR, putative sites for LEF1 are also found in all three species. LEF1 is a factor involved in bone morphogenesis in the Wnt signaling pathway (Westendorf et al. 2004), as well as response to survival signals in pro-B-cells (Busslinger 2004). There are also conserved putative binding sites for ETS-1, which is a proto-oncogene that has been shown to interact with many other transcription factors, one of them being PAX5, to regulate B-cell differentiation and immunoglobulin heavy chain expression (Dittmer 2003). It was therefore of special interest that two PAX5 binding sites were predicted overlapping the transcription-related footprint found on the expressed allele in fibroblast, possibly indicating a complex important for the transcription of *NDN*. PAX5 plays a central role in specifying B-cell identity (Busslinger 2004) and has been shown to be involved in the histone exchange removal of H3K9 methylation during VDJ rearrangement (Johnson et al. 2004). Our data shows that while the highly related Pax2 and Pax8 paired-domains are able to bind *NDN* upstream sequences *in vitro* (Figure 3-4), PAX5 itself is not detectably expressed in our fibroblast lines (Figure 3-5), nor are the sites conserved with rodents. *NDN* is also not expressed in blood lymphocytes, which would include cells of the B-cell lineage. Therefore a link between *NDN* and B-cell function is not supported, but cannot be ruled out in other cell types.

CPBP and *EGR-1* are expressed in the fibroblast cells in which the footprinting assay was performed, and as such, remain candidates for factors binding at the footprint. While the position of their predicted binding sites are not completely overlapping the *in vivo* footprint, their binding and associated complex may change local DNA structure and change sensitivity to DNase I, DMS and UVC modification. CPBP (core promoter binding protein) is a ubiquitous factor that functions in many genes, more specifically, at TATA-less promoters (Koritschoner et al. 1997) like the promoter of *NDN*. EGR-1 is an immediate-early gene that is involved in responses to a variety of signaling cascades that modify transcription of target genes. More specifically EGR-1 has been implicated in synaptic plasticity (Simon et al. 2004) and oncogenesis (Adamson et al. 2005). Whether EGR-1 participates in the same pathways as *NDN* does in brain development remains to be investigated.

Putative binding sites for MOK-2 are conserved between all three species. With respect to the human footprinting data, one of the MOK-2 sites overlaps a DNase I hyposensitivity region on the maternal allele in both fibroblasts and lymphoblasts, making this one of the footprinted regions that is allele-specific regardless of transcription. This special type of footprint may represent occupation by proteins involved in allele discrimination, or remnant of the original imprint from early in development. *MOK-2* is preferentially expressed in brain, testis and transformed cells (Ernoult-Lange et al. 1990). It is also expressed, albeit weakly, in the PWS and AS fibroblasts (Figure 3-7), and is predicted to be expressed in our virus-transformed lymphoblasts. Whether or not MOK-2 is represented by one of the protein-DNA complexes common to fibroblasts and lymphoblasts binding the NDNMOK oligonucleotide in the EMSAs remain to be tested through supershifting experiments with MOK-2 antibodies. MOK-2 functions to repress the retina gene *IRBP* (Arranz et al. 2001), possibly through its association with nuclear lamina (Dreuillet et al. 2002). A similar mechanism of repression of the maternal allele through nuclear matrix association is possible and would be consistent with previous models of imprinted regulation (Greally et al. 1999; Kagotani et al. 2002). The validity of this model as it applies to *NDN* awaits experimental verification by FISH and ChIP techniques.

EMSAs performed with nuclear extracts of the patient fibroblasts and lymphoblasts show the footprints are not merely present because the cell lines differentially produce factors able to bind the *NDN* upstream sequences. More importantly, the presence of factors able to bind the oligonucleotides in cell lines where no footprint is found (Figure 3-6) suggests that the lack of footprints is due to differential accessibility of those sites. This is consistent with differential chromatin states between the two alleles. Patterns of protein-DNA complexes between lymphoblasts and fibroblast extracts are also not strikingly different (Figure 3-6), suggesting tissue-specific lack of footprints may also be in part due to chromatin differences instead of a lack of *trans*-acting factors between fibroblasts and lymphoblasts. Further investigation of the differences in chromatin structure and composition of the two alleles, as presented in the following chapters, will clarify the state of the chromatin at *NDN/Ndn* and how this regulates its allele-specific and tissue-specific transcription.

Chapter 4 ♦ Tissue-specific and imprinted histone modifications of the human *NDN* gene

Parts of this chapter have appeared in:

Lau JC, Hanel ML, Wevrick R (2004) Tissue-specific and imprinted epigenetic modifications of the human *NDN* gene. Nucleic Acids Res 32:3376-3382

Introduction

Chapter 3 explored the regulation of a representative transposed imprinted gene, *NDN/Ndn*, at the level of the promoter. Sequence analysis revealed phylogenetic footprints that correlated with *in vivo* footprints in the promoter region of *NDN* (Figure 3-2). Analysis of potential protein factors binding to these promoter sequences suggested that factors involved in regulating *NDN* had differential access to the two parental alleles. It was then important to investigate further the chromatin states of the two alleles and how this correlates with the observed footprints. To this end, this chapter will describe work on the histone modifications in both the promoter, and the entire transcription unit of *NDN*.

As discussed in the Introduction, epigenetic processes such as DNA methylation and histone modifications are associated with many processes such as transcriptional regulation (El-Osta and Wolffe 2000; Litt et al. 2001; Noma et al. 2001), X-inactivation (Beard et al. 1995; Keohane et al. 1998; Boggs et al. 2002) and genomic imprinting (Brannan and Bartolomei 1999; Mann et al. 2000; Grandjean et al. 2001; Xin et al. 2001). Mechanistic links between DNA methylation and histone modification have been proposed, whereby histone H3 K9 methylation can direct DNA methylation (Rice and Allis 2001; Tamaru and Selker 2001), which can then in turn recruit histone deacetylases, thereby creating a closed chromatin conformation that inhibits transcription (Nan et al. 1998; El-Osta and Wolffe 2000; Jones and Takai 2001). DNA hypermethylation is generally associated with decreased gene activity (Jones and Takai 2001; Reik et al. 2001). Histone hyperacetylation is associated with actively transcribed genes and genes poised for transcription (Razin 1998). Histone H3 can be methylated on lysine 4 and lysine 9 residues to mark either active or inactive chromatin (Jenuwein and Allis 2001; Noma et al. 2001). Whether the same mechanisms operate in tissue-specific control and allele-specific control is less well understood.

The *SNRPN* gene, located in the Prader-Willi Syndrome (PWS) region on 15q11-q13, and the *H19/IGF2* gene pair are among the most intensively studied imprinted genes (Ferguson-Smith and Surani 2001). These genes contain imprinting control elements that control germline imprint resetting of genes located in *cis*, even over large distances (Leighton et al. 1995; Dittrich et al. 1996; Horsthemke 1997). Allelic epigenetic differences found at these imprinted loci with closely associated Imprinting Centers (IC) can therefore be either associated with the IC itself, or be the result of a response to the IC (Pedone et al. 1999; Schweizer et al. 1999; Saitoh and Wada 2000). Current imprinting models do not address mechanisms for the extension of the epigenetic mark to target genes at a distance from their IC, nor the mechanisms for coordinate allele- and tissue-specific expression (Hu et al. 1998; Hanel and Wevrick 2001).

As described in the rationale in Chapter 3, *NDN* serves as an excellent model for the transposed imprinted genes. Our laboratory has previously demonstrated developmentally dynamic patterns of maternal hypermethylation and paternal hypomethylation of the promoter CpG island in mouse *Ndn*, by sodium bisulfite sequencing (Hanel and Wevrick 2001). In this chapter, I will describe characterization of finely mapped regions of histone acetylation and histone methylation surrounding *NDN* using antibody specificities previously shown to be differentially modified in imprinted regions. In contrast to *SNRPN*, *NDN* has a tissue-specific expression pattern and is expressed in brain and fibroblasts among other tissues but is silent in blood lymphocytes and derivative lymphoblastoid cell lines (Jay et al. 1997; MacDonald and Wevrick 1997; Sutcliffe et al. 1997; Hanel and Wevrick 2001). This allows comparisons of DNA methylation and histone modification between tissues in which *NDN* is and is not expressed. The simple intronless genomic structure of *NDN* is also amenable to high resolution ChIP to study the histone modifications over the entire transcription unit. These data were compared with complementary experiments by Dr. Meredith Hanel on the allele-specific methylation of a promoter CpG island and a second downstream CpG island in human *NDN* by sodium bisulfite sequencing. We have evaluated the relative contribution of epigenetic changes associated with tissue-specific gene expression versus those associated with genomic imprinting. Our results suggest that DNA methylation and histone H3 K4 dimethylation and trimethylation epigenetically differentiate alleles in *NDN*, while histone acetylation acts in tissue-specific gene regulation.

Results

Histone modification of NDN.

Chromatin immunoprecipitation (ChIP) was performed to investigate histone modifications across the *NDN* gene. In order to analyze the maternal and paternal alleles in isolation, we again used cell lines derived from AS individuals carrying maternal deletion of 15q11-q13, or PWS individuals carrying a paternal deletion. Primer sets were designed to give high resolution coverage (Figure 4-1). Regions B-G covers the transcription unit of

NDN and 5' CpG island, while regions A and H are several kilobases upstream and downstream respectively. We first analyzed fibroblasts derived from PWS and AS patients. Consistent with previously identified patterns of histone H3 acetylation (H3ac), we observed paternal bias in *NDN* in all regions assayed inside and outside of the transcription unit (Figure 4-3). A similar paternal bias in H4 acetylation was also present (Figure 4-2). While differences in acetylation were present across *NDN*, consistent allelic differences were largest in region B, colocalizing with the promoter where there were greater than four-fold differences between alleles. We then performed similar ChIP analysis with antibodies specific for di- and tri-methylated forms of lysine 4 and lysine 9 of the histone H3 tail (H3K4me2, H3K4me3, H3K9me2, and H3K9me3, of which, H3K9me3 cross reacts with H3K27me3 (Perez-Burgos et al. 2004)) and di-methylated lysine 79 of the histone H3 globular domain (H3K79me2). Consistent paternal bias in H3K4me2 was observed over regions B-E (Figure 4-3). The most striking H3K4me2 difference was seen in region B with an average of greater than seven-fold paternal bias. Using trimethyl specific antibodies, a more restricted pattern of paternal bias in H3K4me3 was seen consistently over region B only. H3K4me3 showed approximately three-fold paternal enrichment with very weak or inconsistent biases elsewhere in the gene. Results characterizing H3K9me2 were quite inconsistent, with qualitative maternal biases in some early experiments (Figure 4-2), but a paternal bias in certain regions in others (Figure 4-4). The H3K79me2 and H3K9me3 (and H3K27me3 by cross reactivity of the antibody) antibodies detected a variable and weak trend towards paternal bias (Figure 4-4).

Unambiguous analysis of this modification awaits commercial availability of more specific antibodies. More detailed investigation of these modifications was not pursued further.

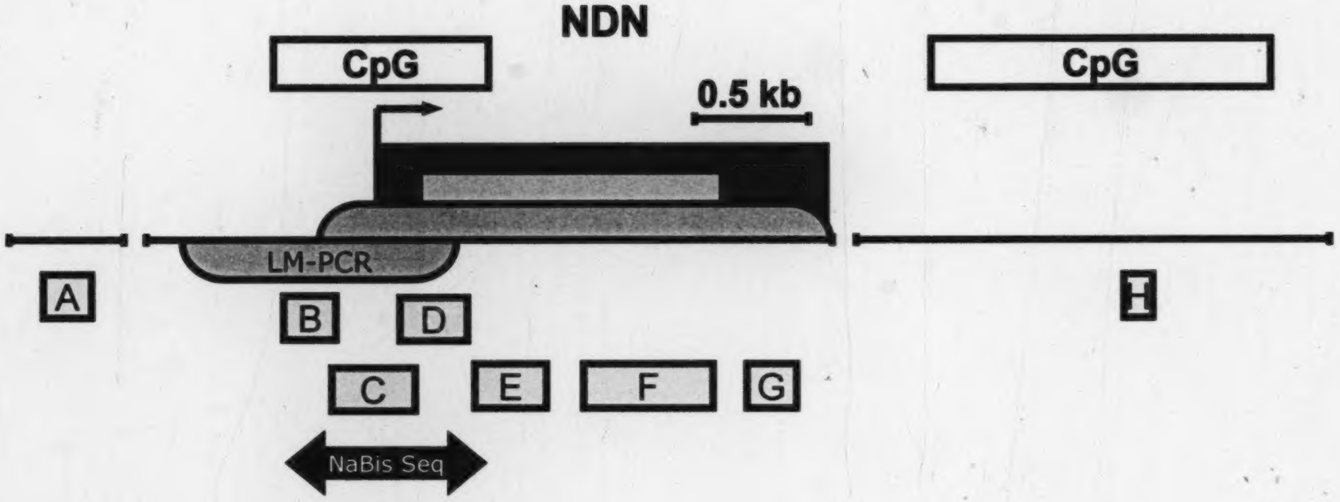


Figure 4-1. ***NDN*** **and surrounding regions.**
Dark green box indicates the single exon of the *NDN* gene, with the ORF indicated by the inset light green box. The arrow indicates transcription start site and CpG islands are as indicated. Regions analyzed with in vivo footprinting in Chapter 3, and bisulfite sequencing by Dr. Meredith Hanel are indicated by "LM-PCR", and double-headed arrows, respectively. Regions analyzed by ChIP are indicated by amplicons A through H. The two gaps each represent approximately 3 kb of DNA. Orange region indicates extent of human-mouse homology.

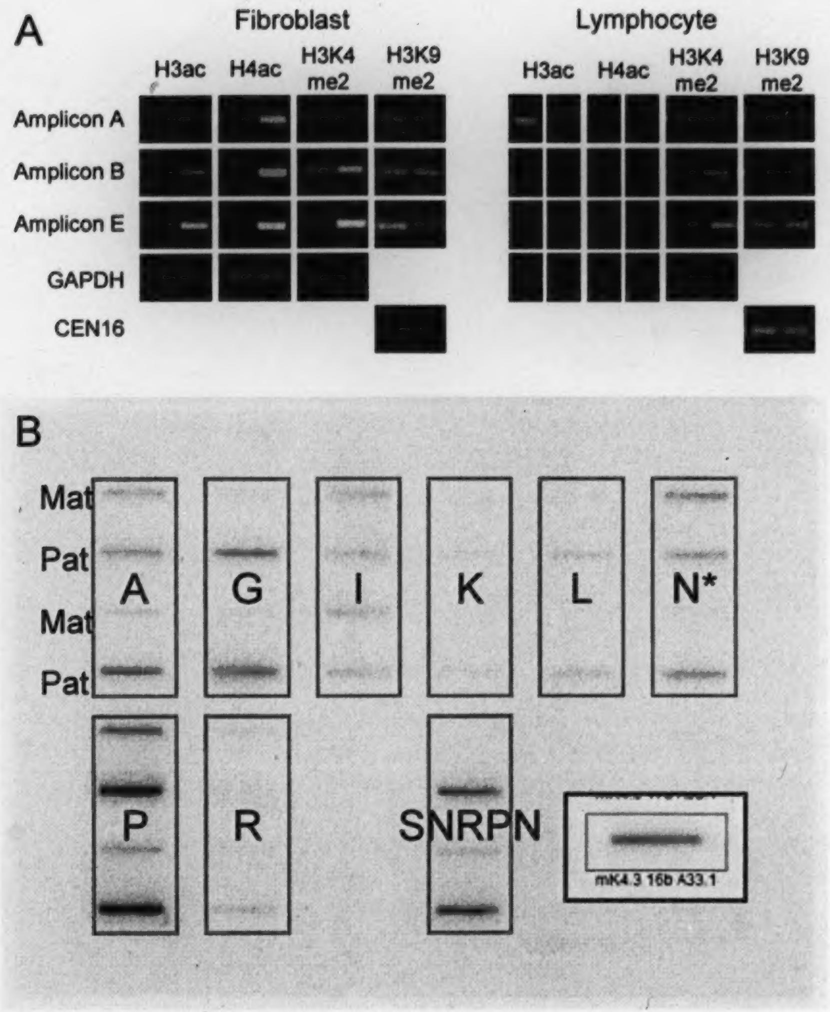

A
Fibroblast
Lymphocyte
H3ac
H4ac
H3K4 me2
H3K9 me2
H3ac
H4ac
H3K4 me2
H3K9 me2
Amplicon A
Amplicon B
Amplicon E
GAPDH
CEN16
B
Mat
Pat
Mat
Pat
A
G
I
K
L
N*
P
R
SNRPN
mK4.3 16b A33.1

Figure 4-2. Examples of qualitative and quantitative ChIP analysis. (A) Qualitative analysis of H3ac, H4ac, H3K4me2, and H3K9me2 in PWS (left band in each frame) and AS (right band in each frame) fibroblast and lymphocytes. Note the apparent maternal bias in H3K9me2 in amplicon E. (B) Quantitative analysis of band intensities by phosphorimager scanning of radiolabeled probes annealed to slot-blotted PCR products. N* indicates a region overlapping with N. Lower right inset is an example of the quantitation object drawn around a band in the ImageQuant software.

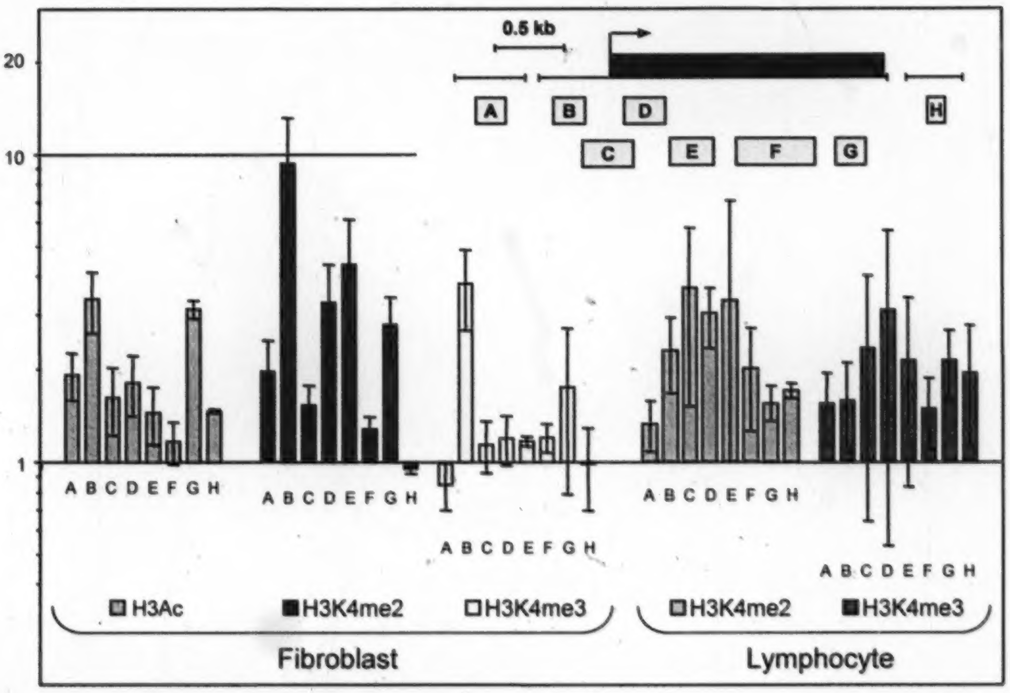


Figure 4-3. Histone modifications of active transcription.
Quantitative ChIP data from experiments using antibodies against H3ac, H3me2K4, and H3me3K4 in PWS and AS fibroblast and lymphocyte data is shown. Paternal to maternal ratio of a representative trial plotted on logarithmic scale where one indicates no bias, greater than one is a paternal bias and less than one is maternal bias. Letters correspond to amplicons assayed as described in (Figure 4-1). Shown is a representative trial, with error bars indicate variation of multiple rounds of detection.

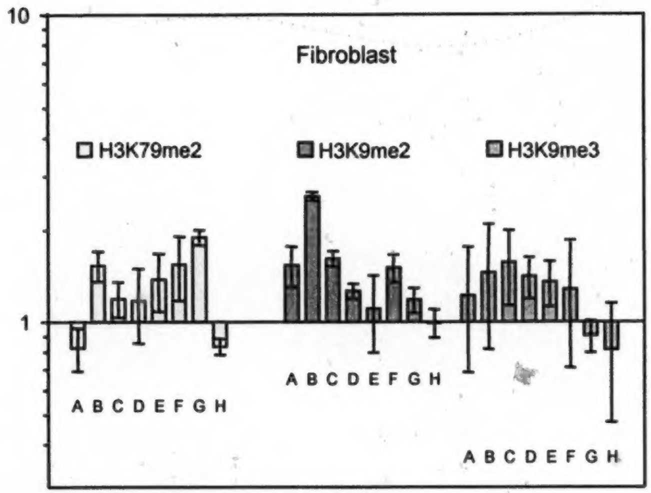


Figure 4-4. Preliminary quantitative data on ChIP in PWS and AS fibroblast lines.
Antibodies against H3K79me2, H3K9me2 and H3K9me3 were used for the experiments below. Note that the paternal bias in H3K9me2 is not always observed (Figure 4-2).

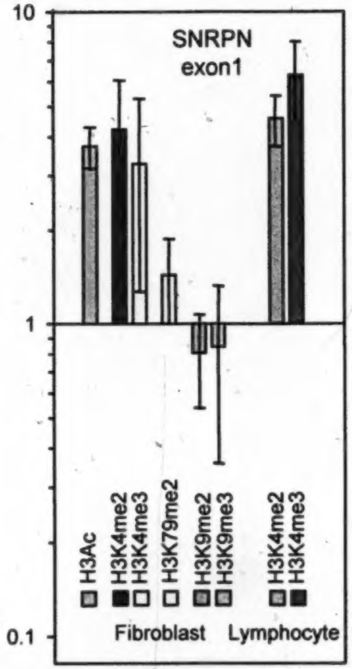


Figure 4-5. ChIP data with antibodies against various histone modifications at SNRPN.
Data presented as in Figure 4-3.

We next performed similar experiments in patient blood lymphocytes to assay whether or not the H3K4me2 and H3K4me3 paternal biases were correlated with tissue type and *NDN* expression. It was previously reported that a region within region F is not associated with allelic histone acetylation in lymphoblasts (Fulmer-Smentek and Francke 2001) and the paternal allele is associated with histone H3K4me2 in region C in blood lymphocytes (Xin et al. 2001). We confirmed this lack of allelic histone acetylation in PWS and AS LCLs and lymphocytes in region F as well as paternal H3K4me2 of region C in patient blood lymphocytes. No other regions in *NDN* had consistent allelic histone acetylation in lymphocytes (Figure 4-2). H3K4me2 allelic differences were distributed over a wider region than previously reported, covering most regions analyzed, although with a weaker bias than seen in fibroblasts (Figure 4-3). A trend towards paternal enrichment for H3K4me3 was also found in lymphocytes, although the degree and distribution of this bias was much more variable. Other modifications were not investigated in lymphocytes because only a limited amount of patient material was available. Overall, these results define a domain of paternal H3K4me3 lying within a domain of paternal H3K4me2 which itself is contained within a large domain of paternal H3ac in fibroblasts, while lymphocytes show a more general allelic bias in H3K4me2 and H3K4me3 without allelic H3ac.

Histone modification of the IC.

To make comparisons between *NDN* and its imprinting center, we studied *SNRPN/IC*, which is expressed in fibroblasts and lymphocytes. We examined histone modification in exon 1 of *SNRPN*, previously described to be paternally enriched for histone H3K4me2 and maternally enriched for histone H3K9me2 in lymphocytes (Xin et al. 2001). In fibroblasts, an H3me2K4 paternal bias was also seen, while we observed maternal bias in H3me2K9 in only some of our trials (Figure 4-5). We next determined if this bias extended to the trimethylated forms, H3K4me3 and H3K9me3. H3K4me3 was found to be paternally enriched at *SNRPN* exon 1 at a level comparable to the enrichment seen with H3K4me2 (Figure 4-3). Using antibodies specific to H3K9me3 however, did not show significant differences between alleles. In blood lymphocytes, we confirmed the paternal bias previously seen in H3K4me2, and discovered an H3K4me3 bias, as is seen in fibroblasts (Figure 4-5). Similar to *NDN*, only weak and inconsistent biases were seen in H3K79me2. These observations are consistent with the fact that *SNRPN/IC* is expressed from the paternal allele in both fibroblasts and lymphocytes.

Discussion

Concurrent with these studies on histone modifications at *NDN* and *SNRPN*, Dr. Meredith Hanel performed bisulfite sequencing to determine the methylation status of *NDN*. She identified a 5' CpG island of 880 bp, containing 73 CpG sites, located in the promoter region of *NDN*, extending from 335 bp upstream of the start codon, and into the open reading frame (Figure 4-1). A second CpG island is located about 4.3 kb downstream of the *NDN* start codon; no equivalent downstream CpG island was found in the mouse sequence for 30 kb downstream of *Ndn* (Genbank #AC027298). Overall, DNA hypermethylation in the 5' CpG island of the maternal allele compared to the paternal allele was observed in both fibroblasts and blood lymphocytes (Lau et al. 2004). The 3' CpG island carried no allelic DNA methylation patterns.

Our studies of histone acetylation are consistent with findings that developmentally regulated genes, such as *NDN*, are usually associated with domains of hyperacetylation (Hebbes et al. 1994; Forsberg et al. 2000) while changes in gene activity in response to stimuli are more frequently associated with localized changes in acetylation (Parekh and Maniatis 1999). In fibroblasts, in which *NDN* is actively transcribed, we identified allelic acetylation differences in a region of at least 10 kb surrounding *NDN*. Intriguingly, the paternal allele is hypoacetylated in the absence of DNA methylation in lymphocytes, suggesting that at this locus histone deacetylases are recruited by factors that are not dependent on DNA methylation, or that DNA methylation is lost after establishment of the hypoacetylated state. Limited studies of the murine *Ndn* promoter, in a region equivalent to human region D (Figure 4-1), indicate that neither allele is acetylated in liver, where necdin is inactive, whereas at least one allele is acetylated in brain, where necdin is expressed (Forsberg et al. 2000). Thus in both human and mouse, the acetylation state of *NDN* may act transiently in transcriptionally competent and transcriptionally active cells, and does not appear to remain as a longer lasting epigenetic imprinting mark.

In *S. cerevisiae*, H3K4me2 has been associated with euchromatic regions of the genome whereas a H3K4me3 state is only seen in actively transcribed genes (Litt et al. 2001; Noma et al. 2001). In light of this association for H3K4me3, it is of great surprise that lymphocytes, not actively transcribing *NDN*, would carry any paternal bias in this modification. We observed a paternal bias in H3K4me2 and H3K4me3 at *NDN* in lymphocytes (Figure 4-3). This shows a striking resemblance to the β-globin cluster in that inactive β-globin genes still carry H3K4me2 and H3K4me3 modifications, which is in contrast to other developmentally regulated genes (Schneider et al. 2004). As those authors suggest, one possible explanation may be related to the long range function of the β-globin LCR, and at this locus, the PWS/AS IC may share similar mechanisms of action. It is possible that the maintenance of the paternal state within the PWS/AS cluster requires all genes on that allele carry certain epigenetic marks regardless of tissue-specific transcriptional status.

Of greater interest are the wide region of paternal H3K4me2 and the nested region of H3K4me3 in fibroblasts. These modifications have been found to be markers of euchromatic regions and transcribed genes respectively (Litt et al. 2001; Noma et al. 2001). Unlike histone acetylation, histone methylation status has been implicated as an early event in chromatin control (Rice and Allis 2001), with histone H3 methylation on residues lysine 4 and lysine 9 reciprocally marking active chromatin and heterochromatin respectively (Jenuwein and Allis 2001). As allelic H3K4me2 and H3K4me3 is also present in lymphocytes (Xin et al. 2001), we propose a model whereby histone H3 methylation at lysine 4 acts to mark allelic differential chromatin states at the *NDN* locus in response to the IC, and that this histone modification represents a persistent somatic mark of the active allele that allows histone acetylation to regulate expression of *NDN* in a tissue-specific manner (Figure 4-6). Interestingly, it has recently been shown that a promoter-restricted distribution of H3K4me2 is a marker of monoallelic genes (Rougeulle et al. 2003). While these authors were not able to distinguish parental alleles, we show here that a similar bias is present in *NDN* on the paternal allele regardless of expression. Restriction of the H3K4me2 modification to near the promoter of this single exon gene is also consistent with their observations. It remains to be seen whether H3K4me3 also display distribution patterns characteristic of imprinted or other monoallelic genes versus biallelic genes. The multiple levels of histone modification in expressing and non-expressing tissues and persistent allelic identity of this imprinted locus may indicate involvement of remodeling complexes implicated in cellular memory. For example, the human trithorax group ALL-1 complex contains HMT activity towards H3K4, HAT activity, as well as chromatin remodeling activity (Nakamura et al. 2002). It will be of great interest to study association of this or other regulatory complexes to maintenance of imprinting at the PWS/AS cluster.

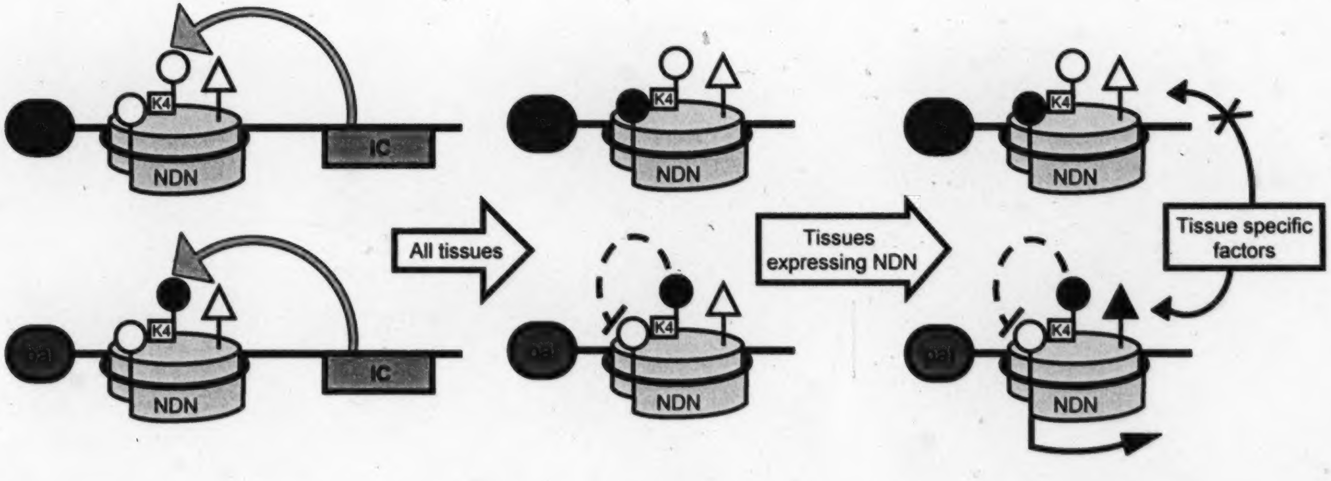


Figure 4-6. Model in which histone methylation plays an early role in imprint establishment and maintenance in NDN.

Pat and Mat refer to alleles of paternal and maternal origin. Black arrow indicates transcription of *NDN* on the paternal alelle in expressing tissue. Lollipops on histone (cylinders) residues and spiral DNA strand indicate allelic biases in histone lysine methylation and DNA CpG methylation respectively. Triangles indicate histone H3 acetylation differences. Solid black symbols indicate bias in allelic modification of that epigenetic mark. The initial signal from the imprinting center can determine histone methylation states, which is translated into other epigenetic marks such as DNA methylation and histone acetylation which sets up a chromatin context for DNA binding factors.

The developmental origins of somatic maternal epigenetic marks are not clear. In mouse and human oocytes, the *NDN/Ndn* promoter is variably methylated, and at least in mouse, differences between the parental alleles are no longer present in blastocysts (El-Maarri et al. 2001; Hanel and Wevrick 2001). Histone H3 K4 methylation on the active allele could serve as a candidate initial epigenetic mark of imprinted target genes, or could translate an initial DNA methylation imprint into a long-term mark that differentiates the two alleles. It remains to be tested whether differential histone H3 methylation exists during early embryogenesis and throughout development and if it acts at the top of a chromatin control hierarchy above allelic DNA methylation and histone acetylation or is simply correlated with these other epigenetic differences. While the mechanism of chromatin changes at target genes by the imprinting center is unknown, our data suggests that allele-specific and tissue-specific epigenetic changes are coordinated for proper gene expression.

Chapter 5 ♦ CTCF binds differentially methylated regions in the imprinted mouse Prader-Willi Syndrome locus

Parts of this chapter have appeared in:

Lau JC, Wevrick R (Submitted) CTCF binds differentially methylated regions in the imprinted mouse Prader-Willi Syndrome locus.

Introduction

Chapter 4 described the histone modification profile in the transcription unit of *NDN*, a representative transposed imprinted gene (Figure 4-3). It is important to study how these modifications apply to the other genes, as well as the conservation of these modifications in other organisms to infer biological significance. Also, since the transposed imprinted genes are co-regulated, it is of interest to investigate mechanisms of domain-wide allelic and tissue-specific regulation of this cluster. The mouse system offers an ideal model to study both these aspects. Mouse chromosome 7C harbors an imprinted domain that is 3.3 Mb in size, and has conserved synteny with human chromosome 15q11-13, the PWS/AS region (Nicholls and Knepper 2001). As described in the Introduction, parental gender of origin-specific gene expression is under the control of the PWS/AS IC located upstream of *SNURF-SNRPN*, which carries a germline maternally DNA methylated region (DMR), paternal histone H3 lysine 4 (H3K4me2) methylation, and maternal H3 lysine 9 (H3K9me2) methylation (Xin et al. 2001). Histone methylation states have often been associated with the regulation of gene activity with methylation of K4 and K9 associated with activity and silence respectively (Peterson and Laniel 2004).

The imprinting of genes located centromeric to the IC on mouse chromosome 7C is mediated by a Megabase long RNA transcript produced from the paternal allele, which controls the paternal allele-specific expression of the contiguous genes *Snurf-Snrpn, snoRNAs,* and also the maternal allele-specific expression of *Ube3a* and *Atp10c* through an antisense mechanism (Chamberlain and Brannan 2001; Runte et al. 2001; Landers et al. 2004). However, there is no indication that a similar RNA-based mechanism maintains imprinting for the contiguous genes *Frat3, Mkrn3, Magel2* and *Ndn,* which are located together on the telomeric side of the IC and all transcribed from the paternal allele. Unlike the *Snrpn-Atp10c* cluster, which are transcribed telomere to centromere for paternal genes and reverse for maternal genes, the paternally expressed *Frat3-Ndn* genes are transcribed centromere to telomere for *Ndn* and *Magel2*, and reverse for *Mkrn3* and *Frat3* (Figure 5-1, top), making an antisense RNA mechanism less plausible. These genes, referred to as the transposed imprinted genes, are thought to have been evolutionarily recent additions to the gene cluster and are proposed to have become imprinted as bystanders (Chai et al. 2001). There is a no homolog for *Frat3* in the human PWS/AS region, where instead the *Mkrn3* homologue *MKRN3* is the most distally located imprinted gene in the PWS/AS cluster.

We therefore investigated other epigenetic mechanisms that could be responsible for maintaining the imprinting status of the transposed imprinted genes. At the *H19/Igf2* imprinted locus, a differentially methylated region (DMR) serves as an epigenetically regulated chromatin insulator to bring about the reciprocal imprinting of these two genes (Bell and Felsenfeld 2000). As discussed in the Introduction, regulation occurs through the differential binding activity of CTCF, a DNA binding protein involved in many aspects of gene regulation. CTCF binds the unmethylated maternal *H19/Igf2* DMR, thereby acting as an allele-specific insulator. Putative CTCF binding sites are also found associated with the *Dlk1-Gtl2* imprinted locus (Takada et al. 2002). To investigate whether CTCF may be involved in the imprinting of the transposed imprinted genes, we first used bioinformatics to predict putative CTCF binding sites. We then verified their *in vivo* binding by PCR-based detection of chromatin-immunoprecipitated DNA (ChIP-PCR), using material from interspecific mouse hybrids to uniquely identify the parental origin of the alleles. We identified two allele-specific CTCF binding sites, one near *Frat3*, the most telomeric imprinted gene in the mouse syntenic region, and a second in the intergenic region between the IC and the transposed imprinted genes. To better understand the epigenetic context of these regions, we investigated histone modifications for the CTCF binding sites as well as their allelic pattern of DNA methylation, and correlated these with the expression patterns of transcripts associated with each site. These results represent the first evidence for CTCF function in the imprinting of PWS region genes.

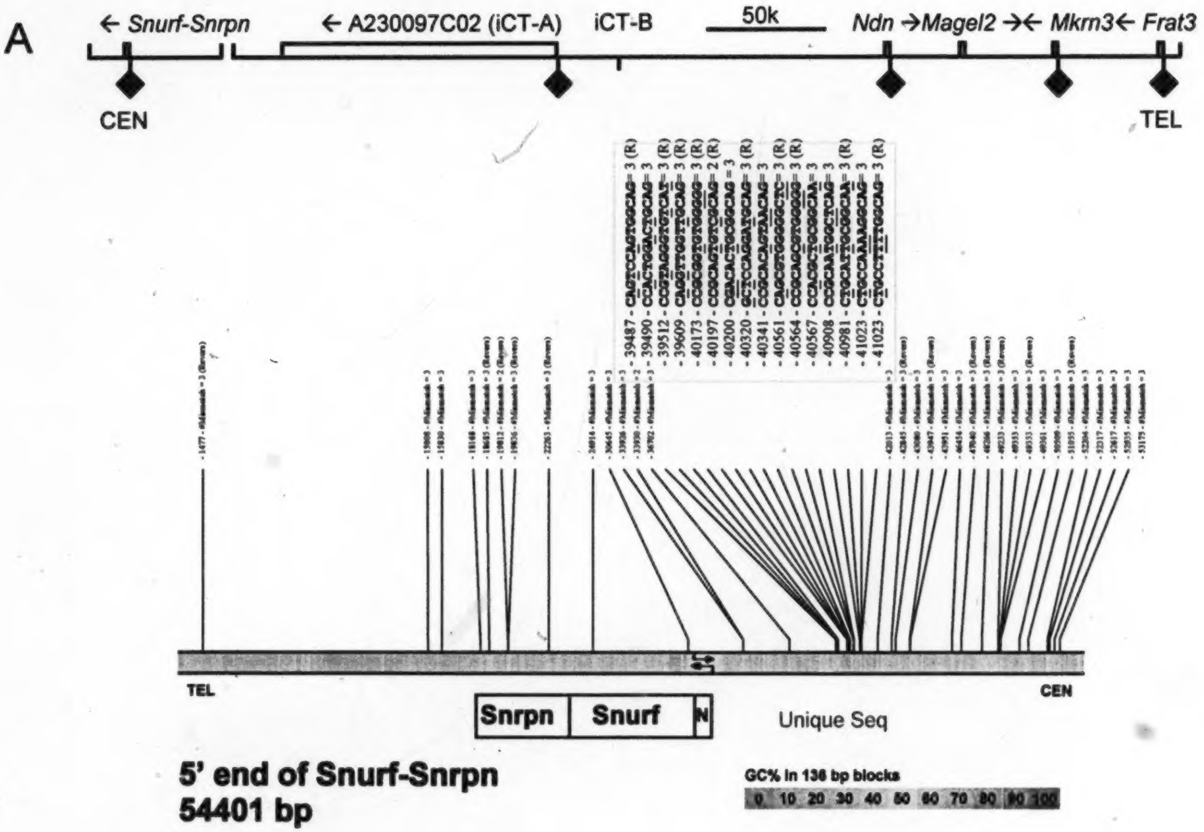
Results

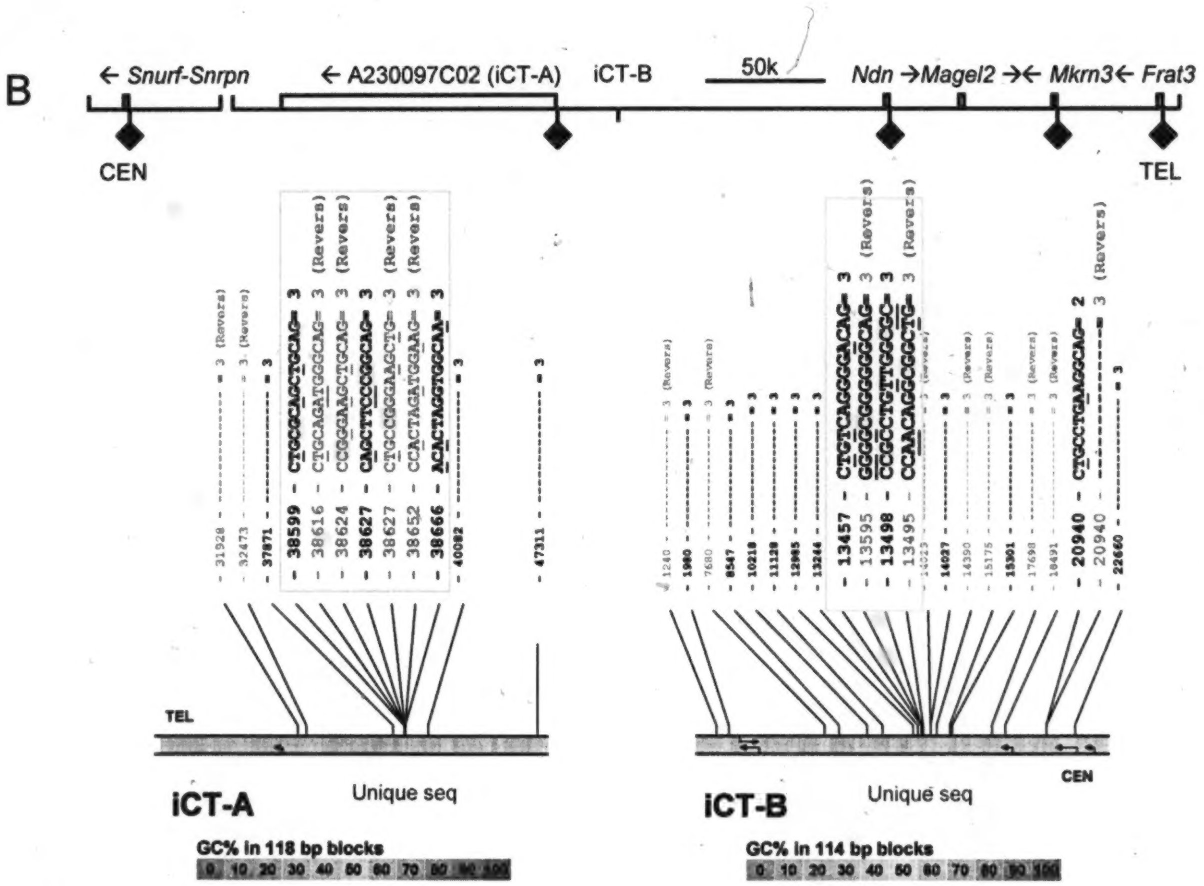
Identification of potential CTCF binding sites in the mouse PWS/AS region

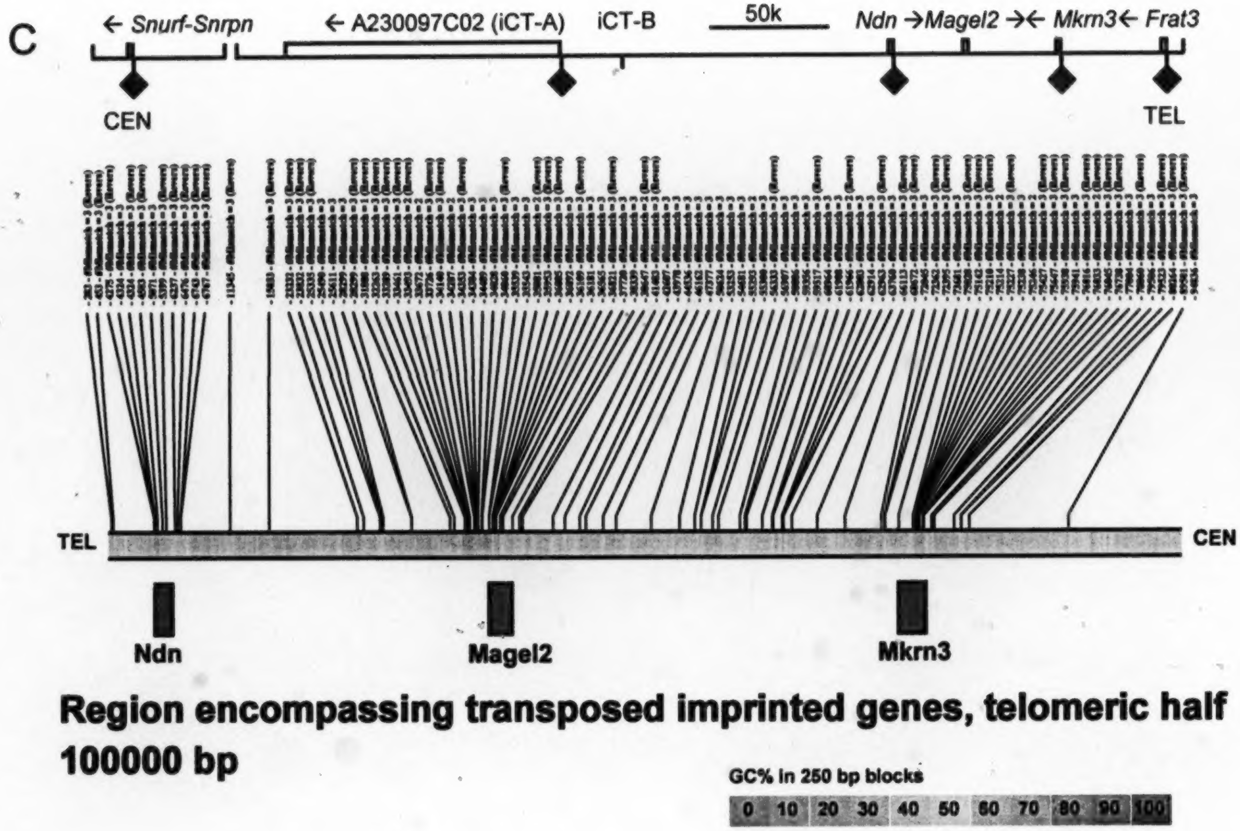
In order to determine a possible role for CTCF in imprinting of PWS/AS genes, we established criteria for the identification of putative CTCF binding sites from primary DNA sequence, and then identified potential CTCF-binding sites within the PWS/AS region. The CTCF protein can bind to a wide variety of DNA sequences in a methylation dependent manner. A CTCF binding site consensus sequence (CCGCNNGGNGGCAG) was previously derived from the *H19/IGF2* imprinted cluster and the X-chromosome (Chao et al. 2002). We used this sequence as our search query, but also allowed for up to three additional single nucleotide deviations from the core consensus sequence in order to identify a wide distribution of possible CTCF-like binding sites. We searched the DNA sequences published in the February 2003 freeze of the UCSC Genome Bioinformatics database (Karolchik et al. 2003). Specifically, we selected a 2.24 Mb genomic region defined as 87 kb centromeric of *Snurf-Snrpn* to 80 kb telomeric of *Frat3*, which included the transposed imprinted genes *Frat3, Mkrn3, Magel2* and *Ndn*. While the sequence was only available in fragments, the vast majority of these fragments were mapped, oriented, and separated by small gaps. As expected, using low stringency search criteria, we found numerous putative CTCF binding sites, totaling 1558 over 2.24 Mb. We then selected regions containing at least four closely spaced CTCF-like sites clustered within 100 bp of each other, which gave a reasonable number of clusters for our initial analysis. We identified multiple clusters of putative CTCF binding sites in the CpG islands of *Ndn*, *Magel2*, *Mkrn3* and *Frat3* (Figure 5-1). There were also putative CTCF binding site clusters in the CpG islands associated with *Snrpn* exon 1 (Shemer et al. 1997). The intergenic region between *Ndn* and *Snrpn*/IC is rich in retroviral repetitive elements, and these repeats contribute to the almost one Megabase difference in the size of this region between mouse and human (~2.24 Mb vs. ~1.5 Mb respectively). Out of 16 CTCF-like clusters in the intergenic region, there were two clusters (intergenic CTCF clusters A and B, hereafter named iCT-A, iCT-B) that were free of repetitive elements. Cluster iCT-A is a CG rich sequence upstream of a predicted mouse gene annotated with the designation of A230097C02, which is supported by multiple spliced ESTs from embryonic head and hypothalamic libraries. The predicted mRNA encodes a putative protein product of 137 amino acids with no significant similarity to known proteins. There is no human counterpart of the A230097C02 transcript in the homologous region. Finally, as predicted by the CG rich nature of the CTCF-like binding core consensus, we identified many CTCF-like binding site clusters in CpG rich regions, including those associated with transposable elements. While these elements may be able to bind CTCF *in vivo*, we did not analyze these further as their repetitive nature prevented unambiguous verification of binding using PCR-based techniques.

Figure 5-1. Predicted CTCF binding clusters.

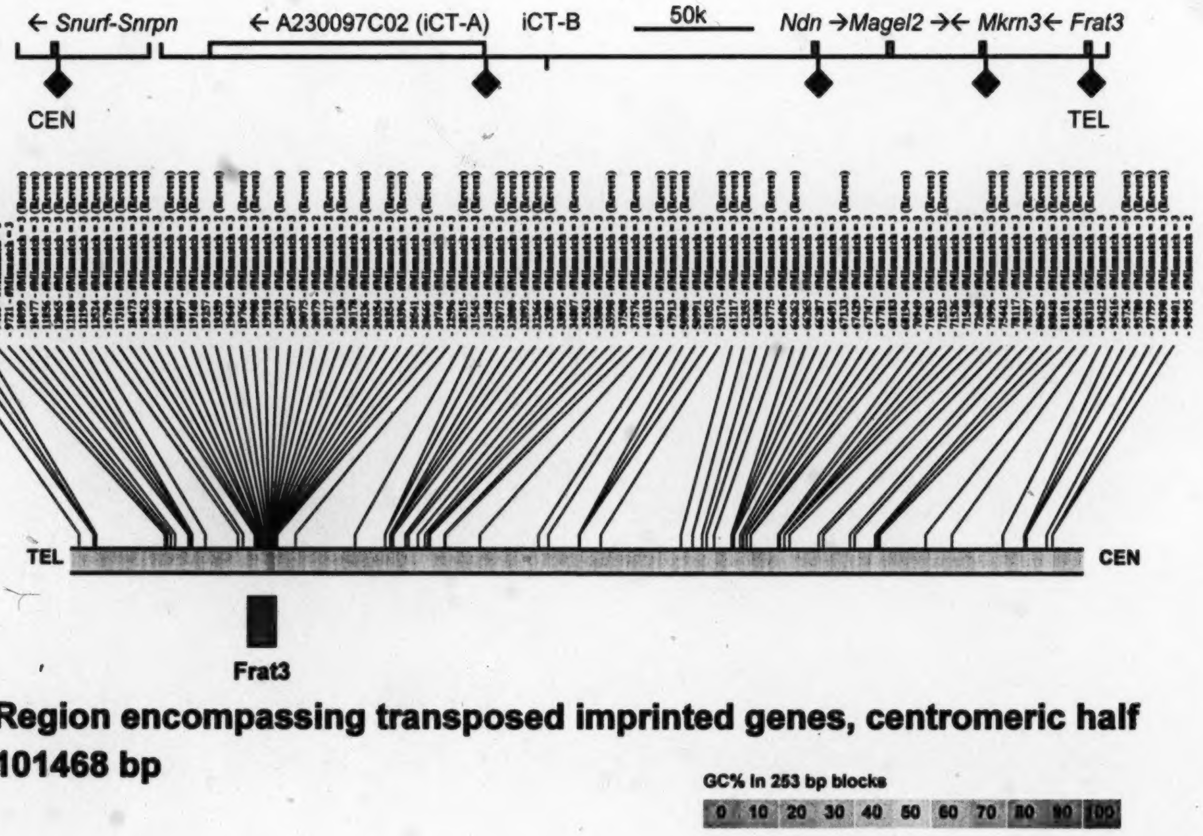
At the top of each figure is a schematic of mouse chromosome 7C, the region of conserved synteny with the human 15q11-q13 PWS imprinted domain. The map is oriented from centromeric (right) to telomeric (left), with a gap between *Snrpn* and A230097C02 of 1.55 Mb. Boxes indicate genes with their name and transcriptional direction above. Diamonds represent CTCF-like binding sequence clusters assayed and their genomic locations. Each figure represents sequence analysis of a section of 7C containing the CTCF clusters of interest. Each vertical line represents one match, with strand and mismatches indicated. CpG dinucleotides are highlighted. (A) region around exon 1 of *Snurf-Snrpn*, (B) iCT-A and iCT-B, (C-D) Region around transposed imprinted genes.


A
← Snurf-Snrpn
← A230097C02 (iCT-A)
iCT-B
50k
Ndn →Magel2 →← Mkrn3← Frat3
CEN
TEL
- 39487 - CAGTCCAGTGGCAG= 3 (R)
- 39490 - CCACTGGACTGCAG= 3
- 39512 - CCGTAGGGTGTCAT= 3 (R)
- 39609 - CAGGTTGGTTGCAG= 3 (R)
- 40173 - CCGCGGTGTGGGGG= 3 (R)
- 40197 - CCGCAGTGTCGCAG= 2 (R)
- 40200 - CGCACACTGCGGCAG = 3
- 40320 - GCTCCAGGATGCAG= 3 (R)
- 40341 - CCGCACAGTAACAG= 3
- 40561 - CAGCGTGGGGGCTC= 3 (R)
- 40564 - CCGCAGCGTGGGGG= 3 (R)
- 40567 - CCACGCTGCGGCAA= 3
- 40908 - CCGCAATGGCTCAG= 3
- 40981 - CTGCATTGCGGCAA= 3 (R)
- 41023 - CTGCCAAAAGGCAG= 3
- 41023 - CTGCCTTTTGGCAG= 3 (R)
TEL
CEN
Snrpn
Snurf
N
Unique Seq
5' end of Snurf-Snrpn
54401 bp
GC% in 136 bp blocks
0 10 20 30 40 50 60 70 80 90 100


B
← Snurf-Snrpn
← A230097C02 (iCT-A)
iCT-B
50k
Ndn →Magel2 →← Mkrn3← Frat3
CEN
TEL
- 31928 - ---------------- = 3 (Revers)
- 32473 - ---------------- = 3 (Revers)
- 37871 - ---------------- = 3
- 38599 - CTGCGCAGCTGCAG= 3
- 38616 - CTGCAGATGGGCAG= 3 (Revers)
- 38624 - CCGGGAAGCTGCAG= 3 (Revers)
- 38627 - CAGCTTCCCGGCAG= 3
- 38627 - CTGCCGGGAAGCTG= 3 (Revers)
- 38652 - CCACTAGATGGAAG= 3 (Revers)
- 38666 - ACACTAGGTGGCAA= 3
- 40082 - ---------------- = 3
- 47311 - ---------------- = 3
TEL
iCT-A
Unique seq
GC% in 118 bp blocks
0 10 20 30 40 50 60 70 80 90 100
- 1240 - ---------------- = 3 (Revers)
- 1980 - ---------------- = 3
- 7680 - ---------------- = 3 (Revers)
- 8547 - ---------------- = 3
- 10218 - ---------------- = 3
- 11128 - ---------------- = 3
- 12985 - ---------------- = 3
- 13244 - ---------------- = 3
- 13457 - CTGTCAGGGGACAG= 3
- 13595 - GGGGCGGGGGGCAG= 3 (Revers)
- 13498 - CCGGCTGTTGGCGC= 3
- 13495 - CCAACAGGCGGCTG= 3 (Revers)
- 14023 - ---------------- = 3 (Revers)
- 14027 - ---------------- = 3
- 14390 - ---------------- = 3 (Revers)
- 15175 - ---------------- = 3 (Revers)
- 15301 - ---------------- = 3
- 17698 - ---------------- = 3 (Revers)
- 18491 - ---------------- = 3 (Revers)
- 20940 - CTGCCTGAAGGCAG= 2
- 20940 - ---------------- = 3 (Revers)
- 22660 - ---------------- = 3
CEN
iCT-B
Unique seq
GC% in 114 bp blocks
0 10 20 30 40 50 60 70 80 90 100


C
← Snurf-Snrpn
← A230097C02 (iCT-A)
iCT-B
50k
Ndn →Magel2 →← Mkrn3← Frat3
CEN
TEL
TEL
CEN
Ndn
Magel2
Mkrn3
Region encompassing transposed imprinted genes, telomeric half
100000 bp
GC% in 250 bp blocks
0
10
20
30
40
50
60
70
80
90
100

D


← Snurf-Snrpn
← A230097C02 (iCT-A)
iCT-B
50k
Ndn →Magel2 →← Mkrn3← Frat3
CEN
TEL
TEL
CEN
Frat3
GC% in 253 bp blocks
0 10 20 30 40 50 60 70 80 90 100

Region encompassing transposed imprinted genes, centromeric half
101468 bp

Verification of CTCF binding in vivo by chromatin immunoprecipitation

To test whether the predicted CTCF binding sites did indeed bind CTCF *in vivo*, we used PCR-based chromatin immunoprecipitation (ChIP) to assay the region surrounding each of the binding site clusters. The CG rich nature of the cluster at the CpG island starting at position -210 upstream of the start codon of *Frat3* precluded consistent PCR based analysis. Therefore a second sequence upstream of the transcript from -781 to -579 was amplified from ChIP DNA. Sequences for PCR amplification in *Mkrn3* and *Ndn* were also chosen 5' to the putative CTCF binding site clusters (-738 to -552 and -582 to +117, respectively). The repetitive nature of the CpG island of *Magel2* precluded PCR analysis. For *Snurf-Snrpn*-IC, a sequence within the CG rich region near exon 1 was analyzed (+60 to +286 with respect to *Snurf* ORF start codon). Regions overlapping predicted CTCF-like clusters were analyzed for iCT-A and iCT-B. The region surrounding the putative CTCF binding site clusters that were chosen for further ChIP analysis were sequenced in DNA from *Mus musculus* (C57BL/6) and *Mus castaneus*. In all cases, we identified polymorphisms that enabled allele-specific analysis in F1 hybrids of crosses between these two species of mice.

ChIP was performed with antibodies against CTCF, using brain and liver tissues obtained from three week old F1 mice generated from a cross between *Mus musculus* (C57BL/6) females and *Mus castaneus* males (Figure 5-2). As a positive control, we tested the DNase I hypersensitive site 1' (HS1') between *TCR* and *Dad1*, which has previously been shown to be highly enriched for CTCF binding compared to a site near HS6 (Magdinier et al. 2004). Both positive control sequences were amplified by PCR from the

DNA obtained by CTCF ChIP, and consistent with published results, we observed a significant difference in CTCF binding between the two sites (Figure 5-2 B). Analogous to findings for the *H19/Igf2* DMR (Bell and Felsenfeld 2000), we predicted that the putative CTCF binding site clusters near the PWS/AS IC might bind CTCF preferentially on the unmethylated paternal allele. Contrary to that expectation, there was no allele-specific binding to the CG rich regions associated with the IC, as both alleles displayed relatively little CTCF binding (Figure 5-2 A). In contrast, the region surrounding the cluster of putative CTCF binding clusters associated with *Frat3* showed strong, statistically significant binding on the paternal allele versus the maternal allele. This result was found in two independent trials and was observed in both brain and liver samples. Furthermore, in the region surrounding the intergenic cluster iCT-A, we observed binding with a paternal bias ($p>0.05$ and $p>0.1$ for the two trials). No significant binding of CTCF to predicted CTCF binding clusters associated with *Ndn* and *Mkrn3* CpG islands was observed on either parental allele. High background levels in the CTCF binding assay at iCT-B precluded quantitative analysis of this site. This is likely due to inefficient probe labeling or annealing, but can also be due to other factors such as probe and primer design.

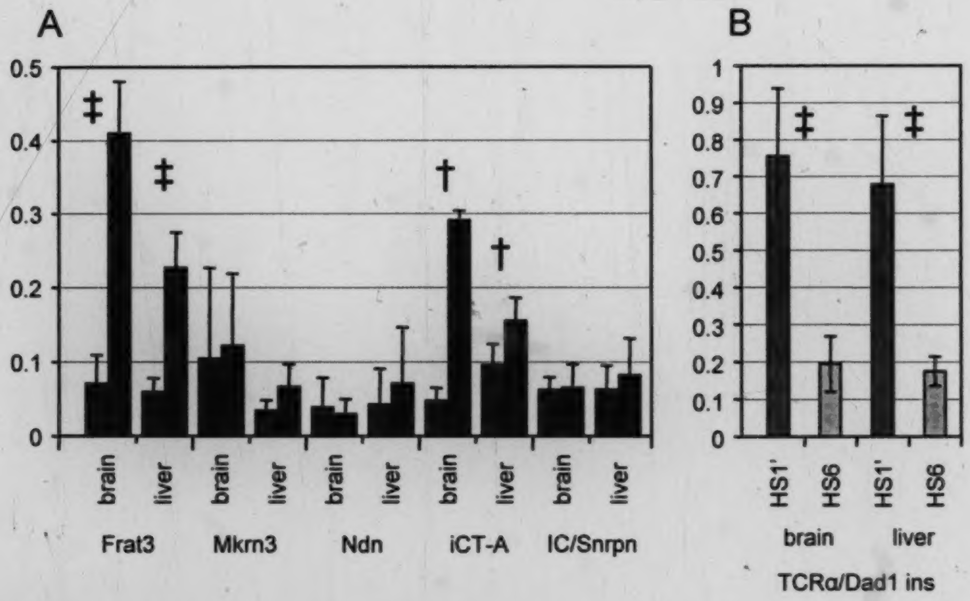


Figure 5-2. CTCF ChIP in mouse telomeric PWS/AS region.
Quantification of allelic differences in CTCF binding at predicted CTCF-like clusters by chromatin immunoprecipitation. Shown here is representative semi-quantitative CTCF ChIP-PCR data. The y-axis indicates binding compared to input, defined as 1% of starting material. (A) The relative amount of binding to the maternal allele (white) and paternal allele (gray) as measured by hybridization of allele-specific oligonucleotides. Double daggers indicate that allelic differences in both trials gave a t-test P-value of less than 0.05, while daggers indicate one trial gave a P-value less than 0.05 and the other less than 0.1. This is interpreted as an allelic bias. (B) Relative amounts of CTCF binding to the intergenic sites between *TCRα* and *Dad1*, performed as a positive control for the ChIP reaction.

Histone modification at putative CTCF cluster sites

To test whether allelic CTCF binding at *Frat3* and iCT-A but not at *Ndn*, *Mkrn3* or the IC was correlated with chromatin context, we used chromatin immunoprecipitation with antibodies that recognize specific histone modifications associated with transcriptional activity. Histone H3 acetylation is typically associated with open active chromatin (Razin 1998). We detected allelic biases in histone acetylation at the *Ndn* and *Snurf-Snrpn* CTCF-like clusters, consistent with the expression patterns and imprinted state of these two genes (Ozcelik et al. 1992; MacDonald and Wevrick 1997). The *Ndn* CTCF cluster had a strong paternal enrichment in histone acetylation ($p>0.05$, both trials, Figure 5-3 A) in the brain where it is expressed from the paternal allele, but not the liver where both alleles are silent. Expression of the *Snurf-Snrpn* gene was detected in both brain and liver, and we observed strong paternal enrichment in histone acetylation in the brain ($p>0.05$, both trials) and a bias towards paternal enrichment in the liver ($p>0.05$ and $p>0.1$ for the two trials), consistent with a previous observation (Fournier et al. 2002). We assessed the expression patterns of *Frat3*, *Mkrn3* and the transcript associated with the intergenic cluster iCT-A (A230097C02) in the brain and liver of 3wk old mice by reverse transcription-PCR (RT-PCR) (Figure 5-4). While expression of *Frat3* was undetectable by RT-PCR from these tissues, the region surrounding the *Frat3* CTCF-like cluster carried strong paternally enriched histone H3 hyperacetylation in brain ($p>0.05$, both trials). *Mkrn3* expression was detected in the brain and not the liver (Figure 5-4), yet it carried strong paternal enrichment for histone H3 acetylation in both tissues ($p>0.05$, both trials, both tissues). The intergenic cluster iCT-A showed strong paternal enrichment in histone H3 acetylation in the brain (p>0.05, both trials), suggesting that the expression of its nearby transcript, A230097C02, in the brain may be paternal allele-specific. Overall, histone acetylation was enriched on the paternal allele of each imprinted gene tested gene, despite the absence of expression of the associated RNA transcript in some cases.

Next, methylation at lysine 4 of histone H3 was examined (Figure 5-3 C-D). Di- and tri-methylation of lysine 4 (H3K4me2 and H3K4me3 respectively) is associated with gene activity in different systems (Sims et al. 2003; Schneider et al. 2004). Consistent with a previous report (Fournier et al. 2002), the CTCF cluster near *Snurf-Snrpn* carried an enrichment of H3K4me2 on the paternal allele (p>0.05, both trials, Figure 5-3 C). At the same site, we also observed a bias towards enrichment of H3K4me3 on the paternal allele (p>0.05 and p>0.1 for the two trials, Figure 5-3 D) in both brain and liver. A bias towards paternal allele enrichment in H3K4me2 (p>0.05 and p>0.1 for the two trials) but not H3K4me3 was observed in *Ndn* in brain only. *Mkrn3* carried a bias towards the paternal allele in H3K4me2 in the brain (p>0.05 and p>0.1 for the two trials), while in the liver, H3K4me2 was paternally enriched (p>0.05, both trials) and H3K4me3 was biased towards the paternal allele (p>0.05 and p>0.1 for the two trials). No significant difference was found for either histone methylation modification in the CTCF binding site cluster near *Frat3*. Finally, the region surrounding the putative CTCF binding sites at iCT-A showed paternal allele enrichment of H3K4me2 and H3K4me3 (p>0.05, both trials, each).

Dimethylation at lysine-79 of H3 (H3K79me2) is associated with developmentally active chromatin regions, and hypomethylation of the same residue has been associated with silencing in yeast (Ng et al. 2003). We found paternal biases in H3K79me2 in the clusters associated with *Ndn* and iCT-A in the brain (p>0.05 and p>0.1 for the two trials of each, Figure 5-3 B), while *Snurf-Snrpn* carried strong paternal enrichment for H3K79me2 (p>0.05, both trials). We investigated di- and tri-methyl states of the K9 residue, associated with silencing in many systems (Jenuwein and Allis 2001), but found either weak or inconsistent binding to most sites (Figure 5-3 E-F). A maternal bias for tri-methylated H3 K9 was seen in the *Ndn* CpG island (p>0.05 and p>0.1 for the two trials).

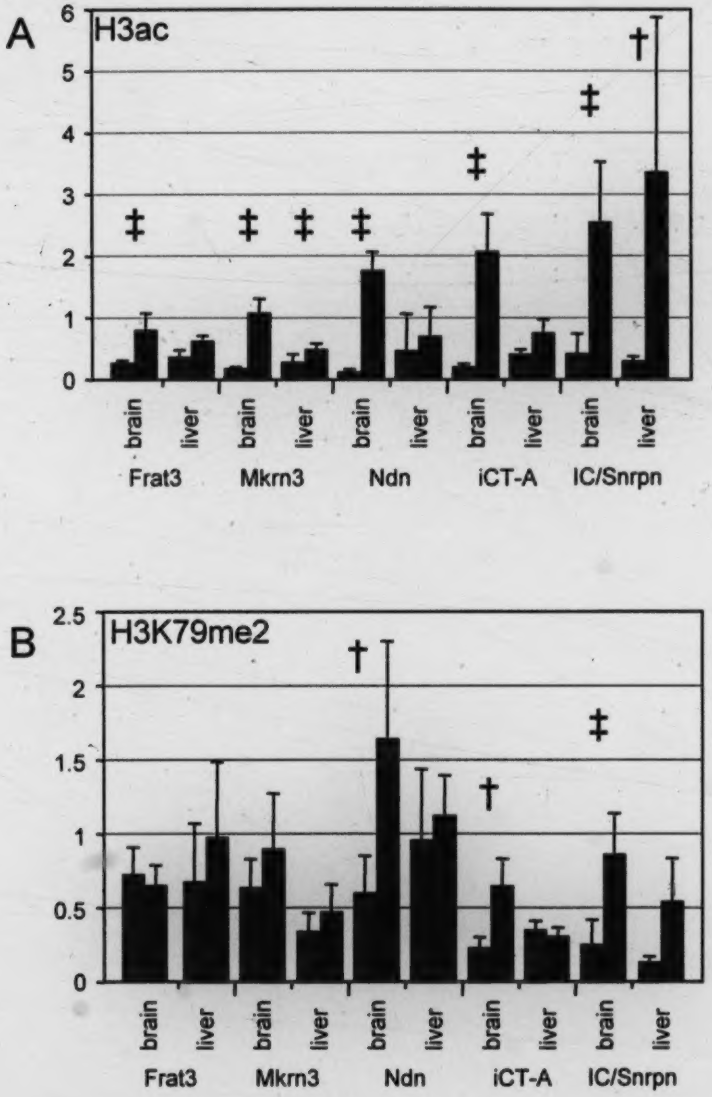
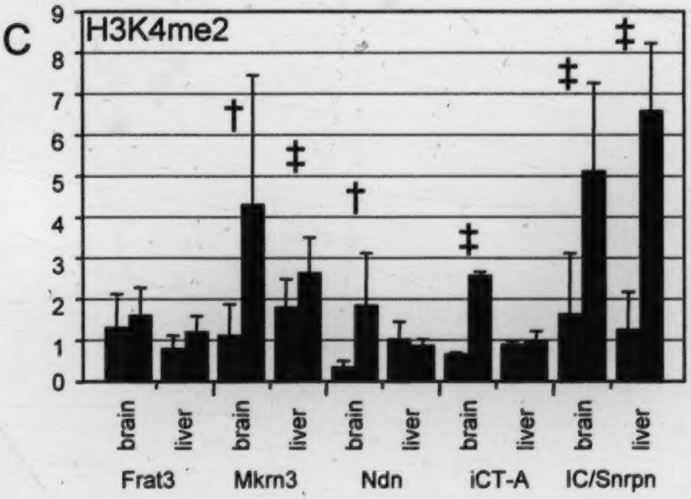
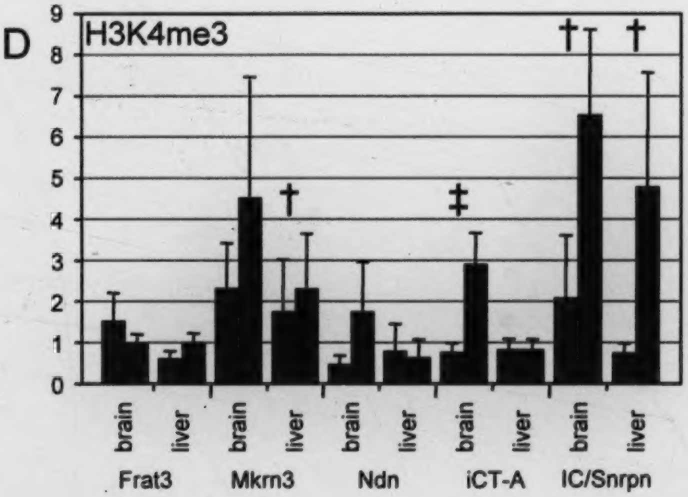
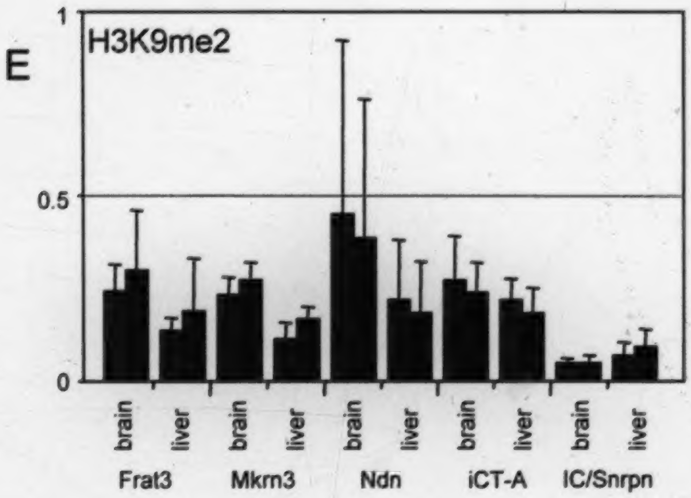
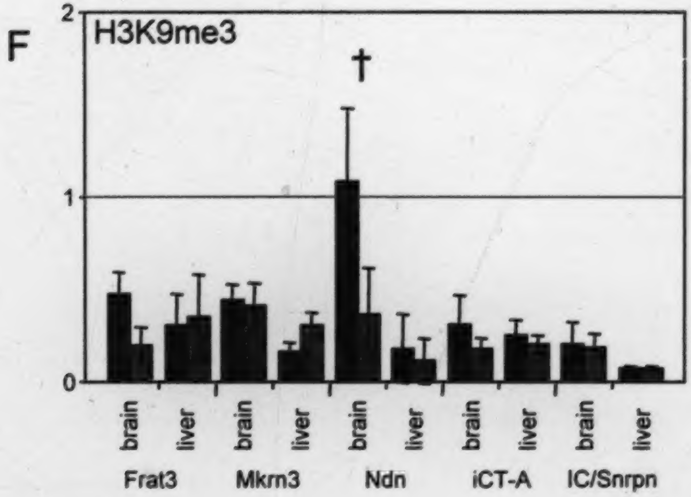


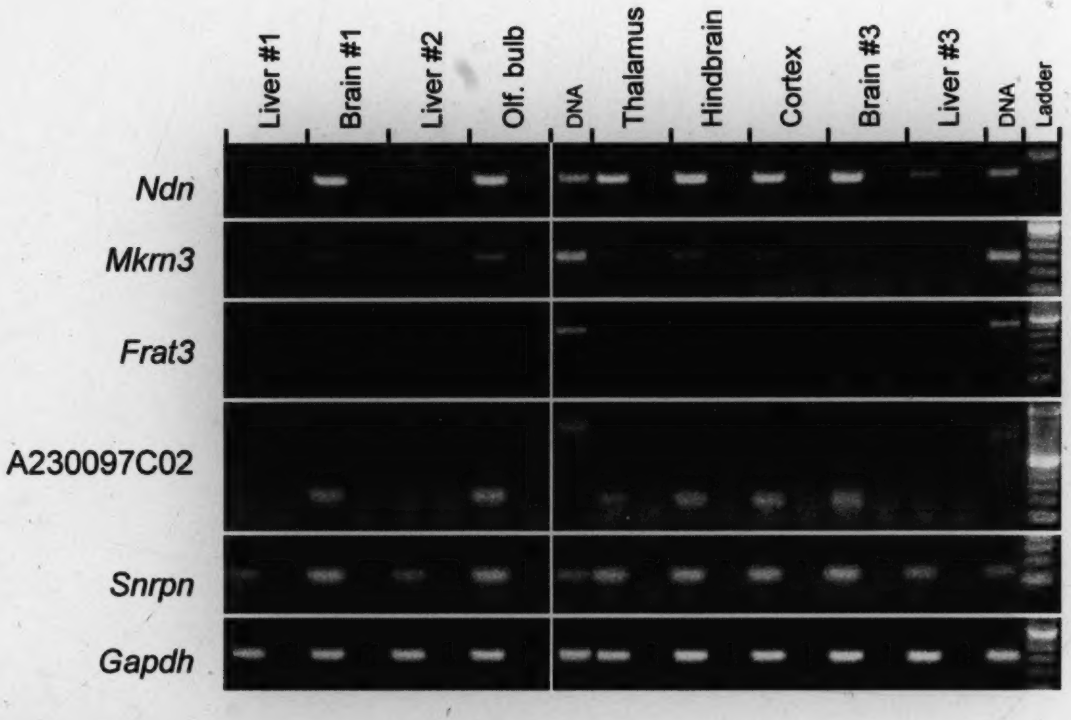
Figure 5-3. Histone modifications at predicted CTCF clusters. Histone modifications of the putative CTCF binding site clusters. Shown here are representative semi-quantitative histone ChIP-PCR data. The data are presented as in (Figure 5-2). (A) Acetylated histone H3, (B) di-methylated lysine 79 of H3, (C) di-methylated lysine 4 of H3, (D) tri-methylated lysine 4 of H3, (E) di-methylated lysine 9 of H3, (F) tri-methylated lysine 9 of H3.


C
H3K4me2
9
8
7
6
5
4
3
2
1
0
†
‡
†
‡
‡
‡
brain
liver
brain
liver
brain
liver
brain
liver
brain
liver
Frat3
Mkrn3
Ndn
iCT-A
IC/Snrpn


D
H3K4me3
9
8
7
6
5
4
3
2
1
0
†
‡
†
†
brain
liver
brain
liver
brain
liver
brain
liver
brain
liver
Frat3
Mkrn3
Ndn
iCT-A
IC/Snrpn


E
H3K9me2
1
0.5
0
brain
liver
brain
liver
brain
liver
brain
liver
brain
liver
Frat3
Mkrn3
Ndn
iCT-A
IC/Snrpn


F
H3K9me3
†
2
1
0
brain
liver
brain
liver
brain
liver
brain
liver
brain
liver
Frat3
Mkrn3
Ndn
iCT-A
IC/Snrpn



Liver #1
Brain #1
Liver #2
Olf. bulb
DNA
Thalamus
Hindbrain
Cortex
Brain #3
Liver #3
DNA
Ladder
Ndn
Mkrn3
Frat3
A230097C02
Snrpn
Gapdh

Figure 5-4. Expression of transposed imprinted genes.
RT-PCR of transposed imprinted genes in 3 week old liver, brain and parts of the brain. For each tissue, the left lane is with reverse transcriptase, while the right lane is without. *Mkrn3* shows different expression patterns in certain parts of the brain. RT-PCR was performed for 30 cycles.

Overall, the CTCF binding cluster iCT-A showed the most striking allelic differences in that it carried paternal enrichment in acetylated H3, H3K4me2, H3K4me3 and a bias in H3K79me2 in the brain, as well as a paternal bias in CTCF binding in both tissues. The paternal allele-specific enrichment and bias of several H3 modifications in spite of the absence of gene expression suggests an allelic differential chromatin structure even in tissues not actively transcribing the transposed imprinted genes. However, no allelic histone H3 modification was perfectly correlated with allelic CTCF binding *in vivo* at any of the tested sites.

DNA methylation patterns of Frat3, Mkrn3 and iCT-A CTCF cluster sites

Binding of CTCF to DNA is prevented by methylation of the target sequence. To test if allelic DNA hypomethylation of the CpG island of *Frat3* is correlated with the allelic pattern of CTCF binding, we performed bisulfite sequencing on DNA extracted from brain and liver of F1 mice. Seventy seven CpG dinucleotides were assayed for methylation. Of these, seventeen are located within ten of the CTCF-like sites in this cluster of thirteen predicted sites (Figure 5-5 A). In both brain and liver, the maternal allele carried moderate methylation (31% and 24% respectively) across all 77 CpG sites, but the paternal allele carried very little methylation (3% and 2% for brain and liver, Figure 5-6 A-D). The CpG dinucleotides within the putative CTCF sites carried 20% and 18% methylation on the maternal allele in brain and liver, and 0.6% and 0% methylation on the paternal allele in brain and liver respectively. Ten CpG dinucleotides were assayed in *Mkrn3* (Figure 5-6 E-H), four of which were located within the putative CTCF binding sites in this cluster of five sites (Figure 5-5 B). The maternal allele of *Mkrn3* in the brain carried much higher levels of methylation (83%) than the paternal allele (7%) in the brain; equivalent methylation in liver was detected for the two alleles. Methylation levels of the CpG dinucleotides within putative CTCF binding sites were found to be similarly high on the maternal allele in brain and both alleles in liver (82%, 78%, and 58% respectively) but low on the paternal allele in brain (8%). Eight CpG dinucleotides were assayed in the iCT-A intergenic cluster (Figure 5-6 I-L), two of which are within four putative CTCF binding sites in this cluster of seven sites (Figure 5-5 C). As in the sites near *Frat3*, DNA methylation of iCT-A was found to be higher on the maternal allele in both brain and liver (55% and 45% respectively) than the paternal allele (0% and 13% in brain and liver). The putative CTCF binding sites on the paternal allele of *Mkrn3* are hypomethylated in the brain but were not found to bind CTCF (Figure 5-6 F), and thus DNA hypomethylation is not sufficient for CTCF binding to these sites. In this case, the CpG dinucleotides within putative CTCF sites had similar levels of methylation to those not in CTCF sites. Overall, all alleles that showed *in vivo* CTCF binding also displayed hypomethylation of the CpG dinucleotides located within the predicted CTCF binding sites on the paternal allele, as well as hypomethylation of sites in the rest of the CpG island, when compared to the maternal allele.

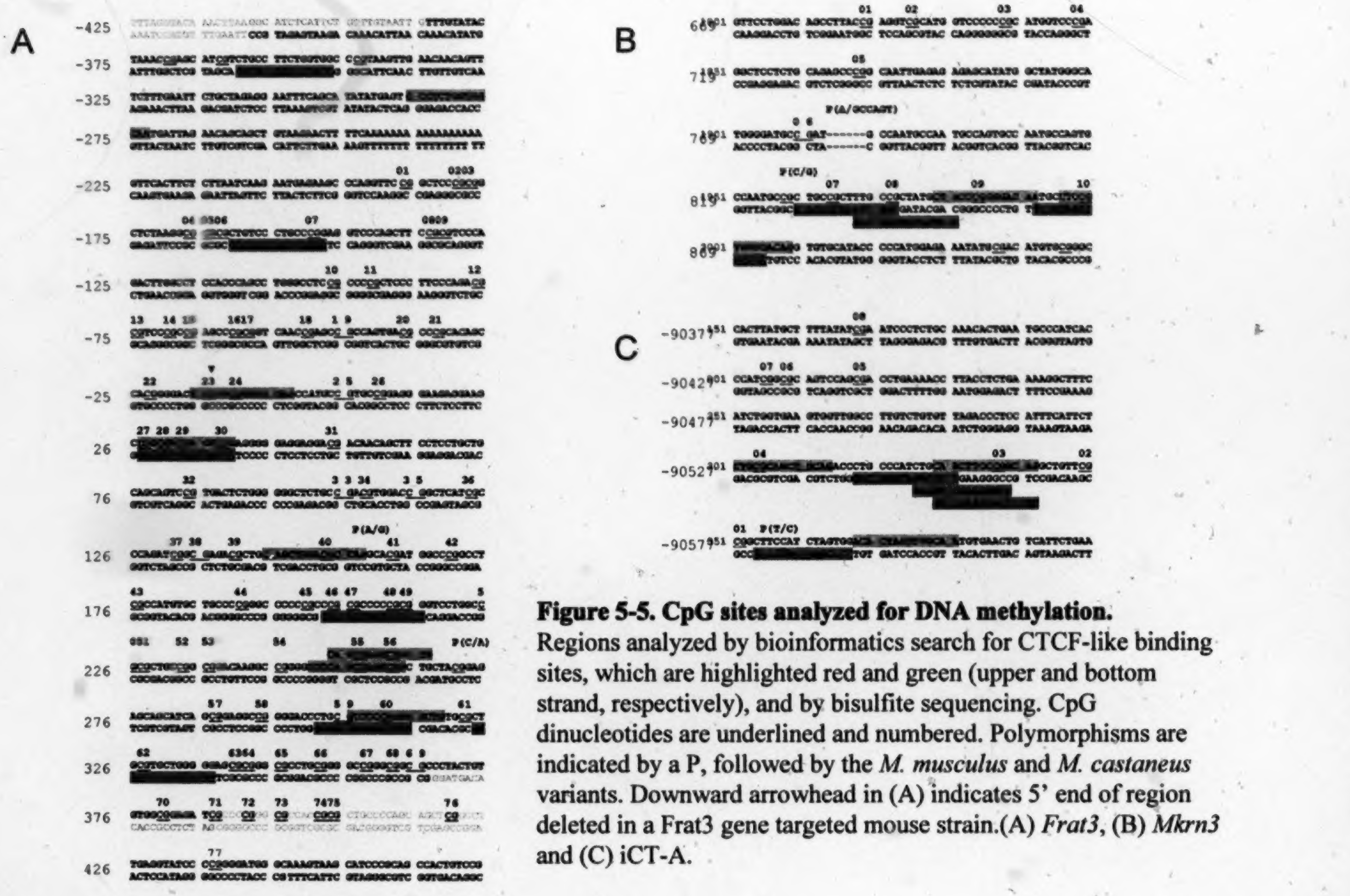

A
-425
-375
-325
-275
-225
-175
-125
-75
-25
26
76
126
176
226
276
326
376
426
B
C

Figure 5-5. CpG sites analyzed for DNA methylation. Regions analyzed by bioinformatics search for CTCF-like binding sites, which are highlighted red and green (upper and bottom strand, respectively), and by bisulfite sequencing. CpG dinucleotides are underlined and numbered. Polymorphisms are indicated by a P, followed by the *M. musculus* and *M. castaneus* variants. Downward arrowhead in (A) indicates 5' end of region deleted in a Frat3 gene targeted mouse strain.(A) *Frat3*, (B) *Mkrn3* and (C) iCT-A.

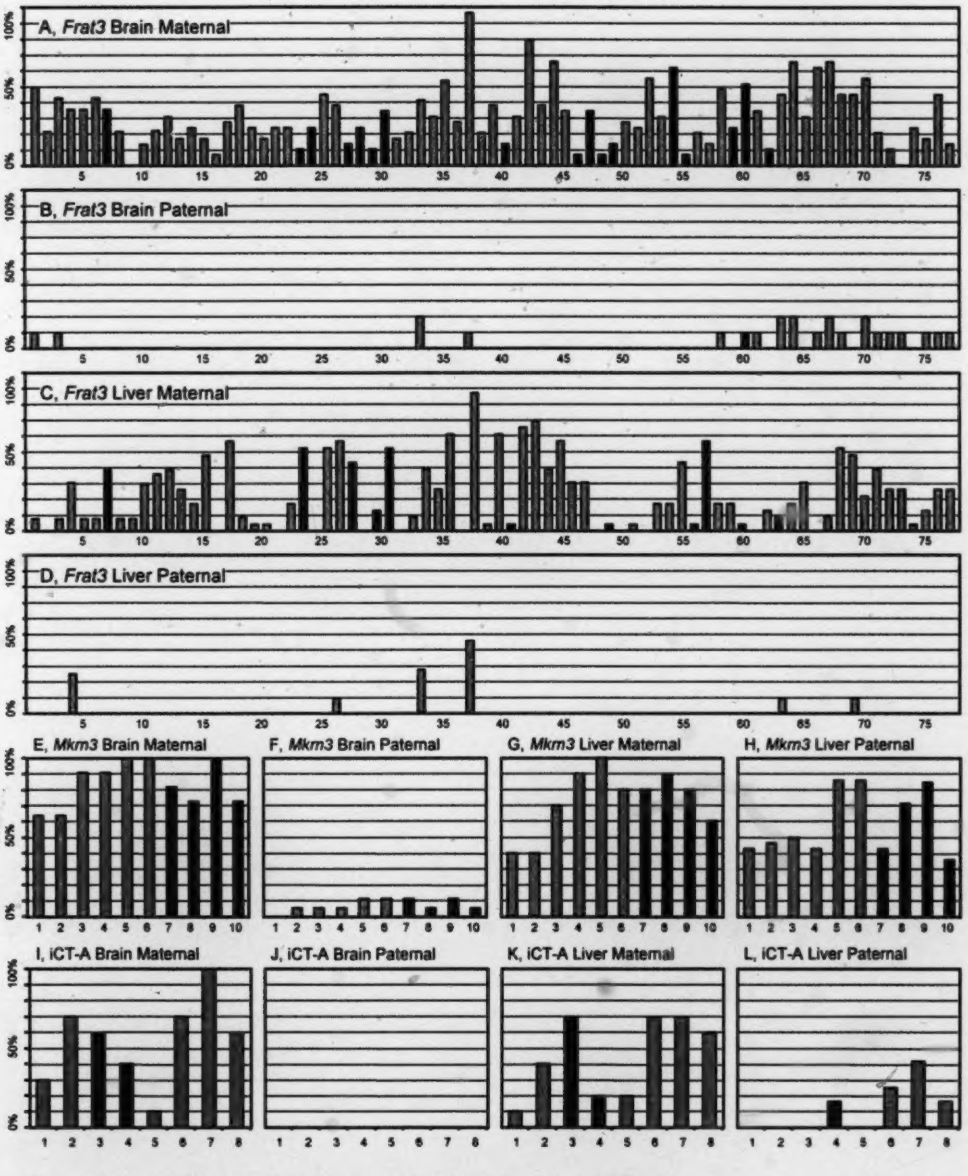


Figure 5-6. DNA methylation of *Frat3*, *Mkrn3* and iCT-A.
DNA methylation patterns on the maternal and paternal alleles of *Frat3, Mkrn3* and iCT-A. The y-axis indicates the percentage of clones that are methylated at each CpG dinucleotide site marked on the x-axis. The bisulfite-treated DNA samples were isolated from brain and liver tissue samples. Sites are numbered as in Figure 1. Black and gray bars indicate CpG dinucleotides within or outside of predicted CTCF binding sites respectively (A-D) *Frat3*. (E-H) *Mkrn3*. (I-L) iCT-A.

Discussion

CTCF binds its target sequence through combinatorial use of eleven zinc finger domains (Klenova et al. 2002). The variation in the use of these domains leads to a considerable variation of CTCF *in vivo* DNA binding sites. The use of specific zinc fingers to bind subsets of CTCF targets also restricts which zinc fingers remain available for interactions with other proteins involved in CTCF functions. Thus, we reasoned that the consensus CTCF binding sequences found at loci where CTCF was involved in mono-allelic gene regulation, such as X-inactivation and imprinting of *H19/IGF2* (Chao et al. 2002), could be shared with CTCF functions at the PWS/AS locus. Using this strategy, we searched the available mouse DNA sequence that has conserved synteny with the human PWS/AS region for a CTCF binding site consensus sequence. This strategy has previously proved useful in finding novel CTCF binding sites at the *H19/Igf2* locus (Ishihara and Sasaki 2002). As this study was designed to test the most likely CTCF binding sites, others may well exist in the PWS/AS region, and this awaits more comprehensive approaches such as those on genomic microarrays. Out of 21 predicted clusters of CTCF binding sites, five were amenable to ChIP-PCR analysis, and we found that only two bound CTCF *in vivo*. Of great interest was the observation that both CTCF binding clusters showed paternal allele-specific CTCF binding. Notably, we also observed that the DNA of the paternal allele was hypomethylated in both tissues studied, correlating with allelic CTCF binding.

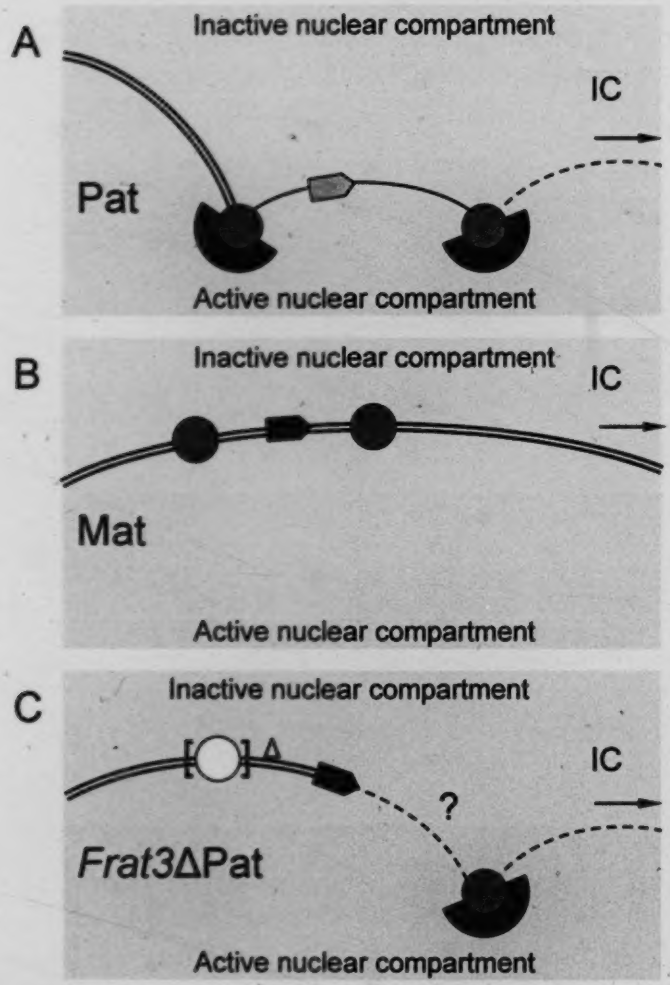
Allelic histone modifications and DNA methylation have previously been shown to be associated with imprinted genes (Delaval and Feil 2004). DMRs are associated with many imprinted genes and some carry germline imprints. Likewise, several histone modifications are found with many imprinted genes (Fournier et al. 2002). Our data suggest no single epigenetic modification is synonymous with imprinted gene regulation and that differences exist between species and even genes in the same cluster. Our previous studies of human *NDN* showed that methylation of H3 lysine 4 was well correlated with allelic identity while histone acetylation was correlated with the imprinted and tissue-specific expression of *NDN* (Lau et al. 2004). In this study, paternal allele-specific H3 lysine 4 methylation and histone acetylation of *Ndn* were both correlated with expression in the brain, while no allelic histone modifications were found in the liver, where *Ndn* is not expressed. Species-specific differences in DNA methylation patterns were also observed. DNA hypermethylation was associated with the maternal allele in human *NDN* regardless of expression (Lau et al. 2004). In the mouse, similar allelic differences were found in the brain, where *Ndn* is expressed in many cells, but both alleles were hypomethylated in the heart, where *Ndn* is not expressed (Hanel and Wevrick 2001). This may be due to selection of representative expressing and non-expressing tissues, since fibroblast and lymphocytes were studied in human and brain and heart were studied in mouse, respectively. There may be either species or tissue-specific differences in the relationship between epigenetic modifications and allele identity. In our current comparison of the transposed imprinted genes in the same cluster as *Ndn*, no definitive epigenetic signature was shared by this cluster of genes. While all genes tested carry allelic histone modifications to some degree in at least one tissue, these modifications do not correlate well with the expression pattern of each individual gene. This may be due to the regions chosen for analysis, since our work

(Lau et al. 2004) and work by others has shown that histone modifications can have very restricted distribution patterns (Schneider et al. 2004). We did observe a general theme of active modifications such as DNA hypomethylation, histone acetylation, and histone methylation on lysine 4 and 79 of H3 on the paternal allele even in some cases where there was absence of expression. This is consistent with a model whereby the paternal allele is more open and carries active modifications, although the repression of transcription depends on tissue-specific factors that act in spite of the permissive modifications at the locus.

Since *Frat3* is not expressed in the adult tissues we tested, it is unlikely that the differential CTCF binding affects transcription, or that the DNA methylation or histone acetylation pattern is related to transcription. Given the critical role CTCF plays in *H19/Igf2* imprinting (Bell and Felsenfeld 2000), and the mechanistic model by which CTCF functions in allele-specific higher-order chromatin structure formation, we propose that the CTCF clusters that we identified also operate to organize chromatin in an allele-specific manner. The germline-derived differentially methylated region (DMR) of *H19* is unmethylated on the maternal allele, binds CTCF, and forms a complex with the upstream maternally unmethylated DMR1 of *Igf2*. This complex forms a loop bringing *Igf2* into an inactive nuclear domain. On the methylated paternal allele of the *H19* DMR, an unknown protein complex is instead formed with the paternally methylated DMR2, at the 3' end of *Igf2*, bringing *Igf2* out of the loop and theoretically giving it access to a nuclear environment permissible to transcription (Murrell et al. 2004). Countless configurations of the PWS region are possible, although one model analogous to the *H19/Igf2* paradigm can be postulated. The *Frat3* CTCF site and the iCT-A site may cooperate to fill the role of a matrix anchor (Yusufzai and Felsenfeld 2004) and bring the transposed genes as a co-regulated unit in and out of active nuclear regions on the paternal and maternal alleles respectively in response to the IC (Figure 5-7 A and B). The *gypsy* insulator of *D. melanogaster* also functions through changes in nuclear localization (Gerasimova et al. 2000). An overall differential chromatin context for the co-regulated unit would also be consistent with our observations that some of the histone modifications we examined were allele-specific without concurrent expression of the associated transcript, in particular *Frat3* itself. It also remains to be seen whether the allelic CTCF binding is conserved in humans as is the case with the *H19/Igf2* gene pair.

Figure 5-7. Model of CTCF function at transposed imprinted genes.

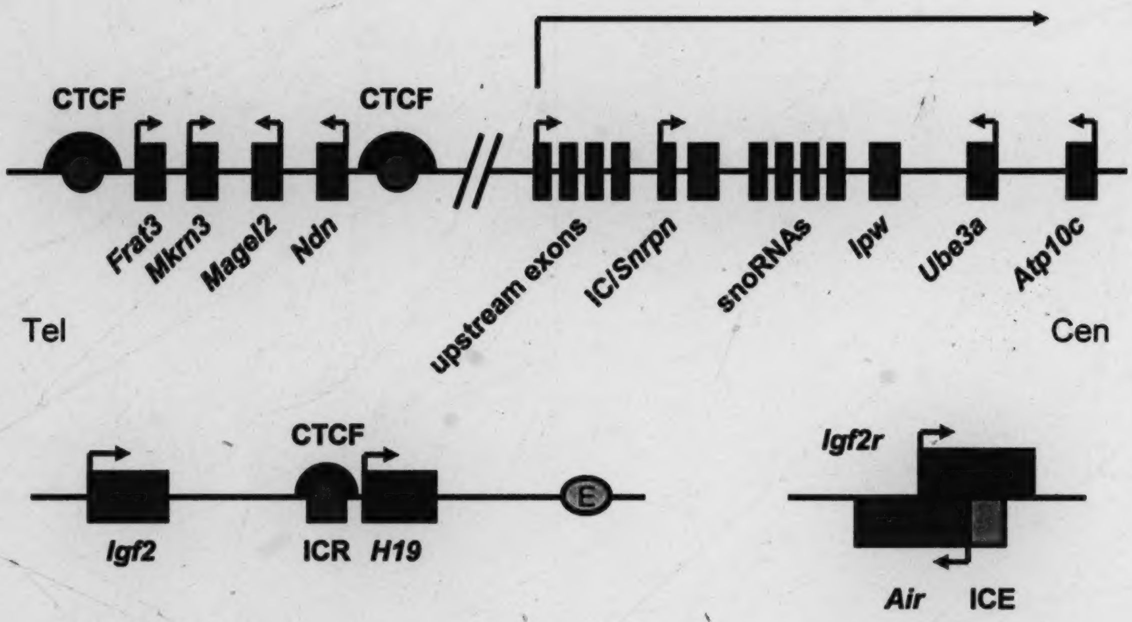
Model for the function of CTCF binding in the transposed genes of PWS/AS 7C imprinted domain. (A) On the paternal allele, flanking CTCF binding sites (filled circles) are bound by CTCF (half donuts) and the intervening transposed genes are sequestered into nuclear compartments permissible to gene activity, thus allowing a euchromatic state (single line). The expression of intervening genes and the state of histone modifications and DNA methylation are dependent on species- and tissue-specific factors. (B) On the maternal allele, the CTCF binding sites are not bound by CTCF, allowing heterochromatinization of the transposed genes along with the rest of the imprinted domain on that allele. (C) Upon deletion of the telomeric CTCF binding site of *Frat3* (empty circle), the paternal allele may partially lose association with active nuclear compartments, allowing position effects of chromatin outside the domain (double line) to encroach. The degree of this effect remains to be tested in this deletion mouse.



A
Inactive nuclear compartment
IC
Pat
Active nuclear compartment
B
Inactive nuclear compartment
IC
Mat
Active nuclear compartment
C
Inactive nuclear compartment
Δ
IC
?
Frat3ΔPat
Active nuclear compartment

Whether the CTCF binding sites identified in this study participate in intralocus association within the PWS region in a manner analogous to the *H19/Igf2* locus await further study, with these two sites being ideal starting points for the chromatin conformation capture assay (Dekker et al. 2002). The CTCF binding site at *Frat3* may have an alternate role in demarcating the telomeric end of the imprinted domain at mouse 7C and may insulate the domain from more telomeric genes, or conversely, may protect other genes from the effects of the long range actions of the IC. Targeted deletion of the CTCF binding sites associated with *Frat3* and the iCT-A clusters would test the hypothesis that these are indeed important in the maintenance of the imprinted states for all four transposed imprinted genes. Three independent gene targeting experiments that modified *Ndn*, *Frat3*, and the highly related but non-imprinted *Frat1* gene are useful in elucidating the possible importance of the distal CTCF binding site cluster in imprinting. In one *Ndn* gene-targeted line, 33% of the CpG island and most of the open reading frame of *Ndn* was replaced with *LacZ*, which became imprinted under control of the *Ndn* promoter (Gerard et al. 1999). A gene-targeted deletion of *Frat3* that includes most of our predicted CTCF-like sites was recently reported (van Amerongen et al. 2005). In this case, the *Frat3* open reading frame, which contains most of the CTCF-like sites, was replaced with a *LacZ* reporter gene. This removed 80% of the CpG island and ten of thirteen CTCF-like binding sites. Surprisingly, the reporter became silent and methylated on both alleles. *Frat1* is the highly homologous ancestral gene to *Frat3*, and is located in a separate region of the genome and is therefore not imprinted. Gene-targeting of *Frat1* removed its open reading frame, coincident with 77% of the CpG island, and allowed correct spatial-temporal embryonic expression of the inserted reporter gene (Jonkers et al. 1999). While the silencing of *Frat3* could be due to removal of basal promoter elements of *Frat3*, the fact that a very similar *Frat1* deletion still enabled correct expression of a reporter gene argues against this hypothesis. We favor the interpretation that the *Frat3* deletion removed CTCF binding sites critical for activity on the paternal allele, which normally protects *Frat3* from position effects telomeric to the imprinted domain (Figure 5-7 C). A similar role for CTCF sites has been proposed for the delimitation of the imprinted *Igf2/H19* and non-imprinted β-globin domains (Saitoh et al. 2000; Ishihara and Sasaki 2002). Functional insulator assays and matrix enrichment assays of this CTCF binding site will help to refine this model. Further examination of expression and epigenetic marks of the other three transposed genes in the *Frat3* transgenic mouse model could be instrumental in understanding the role of the *Frat3* CTCF binding site in genomic imprinting.

The possible role of CTCF in imprinted regulation and maintenance of the PWS domain has been poorly defined. Discovery of the first evidence of CTCF involvement within the PWS imprinted region was made possible by our strategy based on clustering of multiple CTCF-like consensus binding sites. This serves as a critical first step in a more comprehensive documentation of CTCF binding and function in imprinting of PWS genes as well as other large imprinted clusters. Our results here also raise the possibility that the transposed imprinted genes may be regulated through a mechanism similar to the *H19/Igf2* locus, while the genes centromeric (in the mouse) to the PWS IC are imprinted *via* a mechanism involving transcription of an antisense RNA (Chamberlain and Brannan 2001; Runte et al. 2001) in a manner similar to that proposed for the *Igf2r/Air* locus (Sleutels et al. 2002; Landers et al. 2004) (Figure 5-8). The mechanisms whereby the PWS/AS IC can bring about imprinting by more than one mechanism merit further study. With detailed characterization, this unique feature of the PWS/AS imprinted cluster may begin to reconcile seemingly disparate models of imprinted regulation at different loci.


CTCF
CTCF
Frat3
Mkrn3
Magel2
Ndn
upstream exons
IC/Snrpn
snoRNAs
Ipw
Ube3a
Atp10c
Tel
Cen
CTCF
Igf2
ICR
H19
E
Igf2r
Air
ICE

CTCF imprinting mechanism

antisense imprinting mechanism

Figure 5-8. Models of imprinting operating at PWS/AS domain.
Comparison of different models of imprinted regulation. The interpretation of the data presented here raises the possibility that while the centromeric side of the PWS cluster may employ the antisense suppression mechanism similar to X-inactivation and other imprinted clusters such as the *Igf2r/Air* locus, the telomeric side may make use of allelically regulated CTCF binding similar to the *Igf2/H19* locus.

Chapter 6 ♦ Conclusions and Open Questions

The phenomenon of genomic imprinting has enjoyed the attention of many investigators ever since its discovery. By breaking from Mendel's rules, imprinting has become one of the last great mysteries in the study of human heredity. As well as being an irresistible biological puzzle, it also has relevance to human congenital disease, development, oncogenesis and evolution. Imprinting also shares many aspects of gene regulation with other systems such as X-inactivation and regulation of the well studied β-globin cluster. Therefore, detailed investigation of genomic imprinting will give insight into mechanisms applicable to general gene regulation, and will better define the role of imprinting in the scheme of human biology.

The PWS/AS cluster is an excellent example of an imprinted locus. There are well defined human disorders based on isolated deficiency of both alleles, and mutations defining the imprinting center that carries the germline imprint. A well conserved region of synteny in the mouse presents a convenient model organism for genetic dissection and evolutionary conservation. Evidence suggests genes centromeric and telomeric to the IC are controlled by different mechanisms. While a partial model exists for the mechanism of imprinting for one side of the locus (telomeric in human, centromeric in mouse), little is known about the regulation of the transposed imprinted genes. A well-defined imprinted gene, *NDN/Ndn*, was used to investigate the less defined aspects of imprinting at this locus. By studying this gene from varying perspectives, starting with fine scale examination of its promoter, then histone composition for a region encompassing the gene, to domain wide regulation, these studies have given surprising insight into imprinting of the transposed imprinted genes in the PWS region.

Regulation of the transposed imprinted genes

Detailed bioinformatic examination of the promoter sequence of *NDN/Ndn* from human, mouse and rat has shown conservation of promoter sequences, as well as putative binding sites for several transcription factors. Certain features such as conservation in sequence of part of the minimal promoter and predicted sites for factors involved in the control of genes in neuronal function are consistent with some proposed roles for *NDN/Ndn*, since there are overlapping expression patterns of human and mouse in the nervous system. Comparison of this phylogenetic footprint to the *in vivo* footprint information gathered by Dr. Meredith Hanel has been useful in suggesting avenues of investigation on the identity of the *trans*-acting factors regulating *NDN*. While the candidate proteins tested here have not been proven to regulate *NDN* at its promoter, other candidates identified in these studies may prove themselves functional in the future. More importantly, the data presented in Chapter 3 has suggested chromatin accessibility, instead of a purely *trans*-factor based scheme, is also involved in the tissue and allelic regulation of *NDN*.

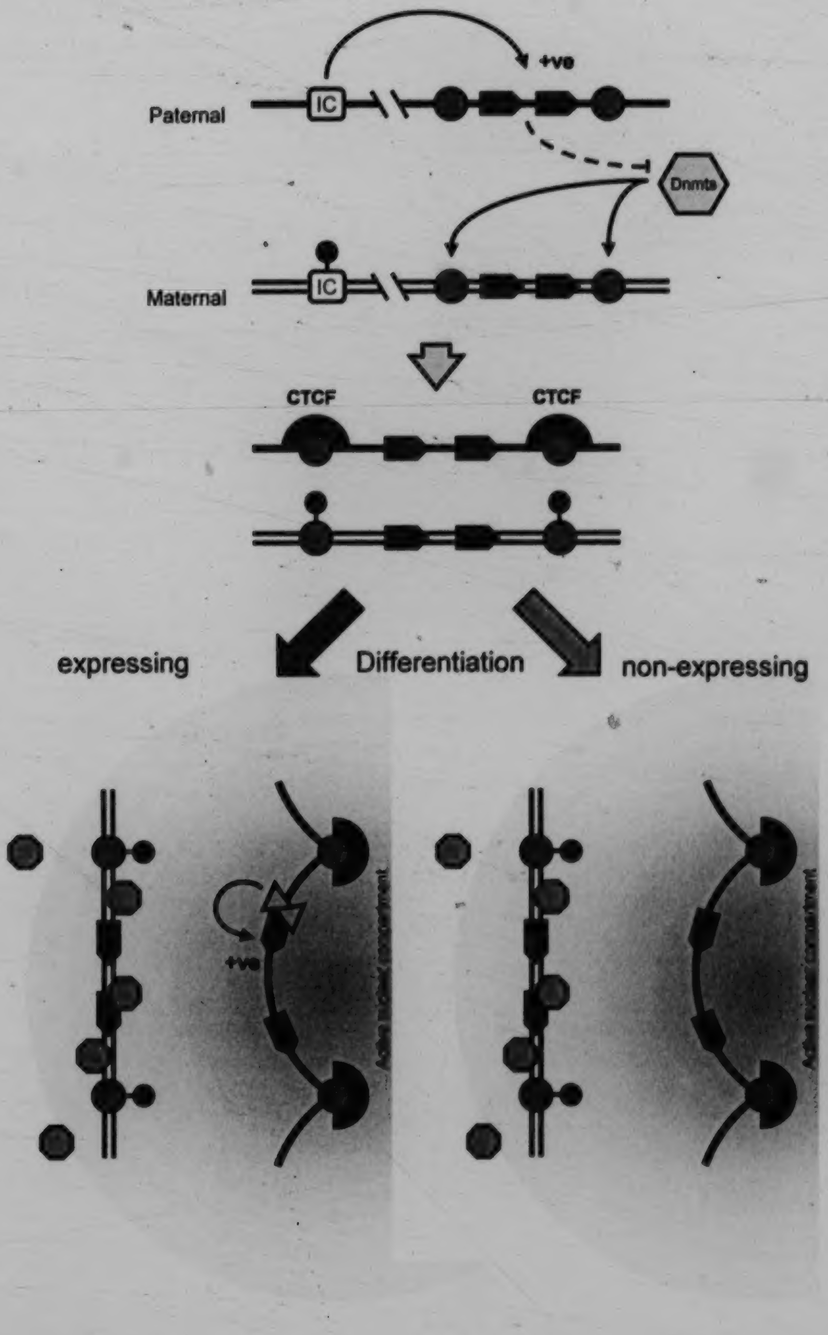
To further characterize differences between parental alleles and expressing and non-expressing alleles, the chromatin context in terms of histone modifications was defined for *NDN* in Chapter 4. As expected, certain modifications known to be associated with activity, such as histone acetylation and methylation at lysine 4 of H3, are associated with the paternal allele in an expressing cell type. In this instance, the modifications also displayed a distinctive distribution where acetylation was spread out over a wide area including sequences upstream and downstream, while dimethylation of lysine 4 was more restricted and trimethylation was only found in a small region within the region of dimethylation. Surprisingly, methylation of lysine 4 of H3 was also found to be paternal allele-specific in the non-expressing cell type tested, although with a less distinct distribution. Therefore, there seem to be different types of "active" chromatin at *NDN*. The paternal allele can carry the active modification of lysine 4 methylation and DNA hypomethylation, whereas in a cell type where *NDN* is also transcribed, acetylation also occurs. This may indicate that lysine 4 methylation is a persistent allelic mark regardless of expression. This is correlated to the situation in the *in vivo* footprint, where the transcription-related footprint is associated with the paternal allele in the expressing cell type, but other footprints are allelically associated in both cell types. Of course the cause and effect relationship between binding of protein factors and histone modifications cannot be determined in these assays. The hierarchical dominance of one over the other is also not mutually exclusive. One scenario can be envisioned where histone lysine 4 methylation represents a persistent mark of the paternal allele, transmitted through cell divisions, and may allow binding of certain factors, some in every cell, and some in cells expressing *NDN*. The binding of factors up-regulating *NDN* in expressing cells may recruit acetylases and chromatin remodelers to aid in opening up the locus and allow binding of other factors and transcriptional machinery. Identification of the factors involved in regulating *NDN* will help define this relationship. Further characterization of histone modifications in other tissue types and developmental stages will also be useful in determining the order of events in the regulation of *NDN*.

To gain a better understanding of the mechanism involved in the differential chromatin states of the two alleles, it was necessary to take a step back and look at the big picture. Since CTCF had previously been shown to have a role in imprinting and long distance gene regulation (Kato and Sasaki 2005), its involvement in the imprinting and regulation of the transposed imprinted genes in the PWS/AS cluster was tested. A bioinformatics search for binding sites similar to previously characterized sites revealed candidate regions where CTCF was predicted to bind. Verification of binding *in vivo* made possible a novel discovery of a possible regulation model for these genes. Two sites flanking the transposed imprinted genes were found to bind CTCF, and did so only on the paternal allele. These sites may delimit a co-regulated domain containing this set of genes and aid in their imprinted regulation. Again, cause and effect cannot be determined at this point, and whether CTCF causes allele-specific organization or differential chromatin states allows CTCF binding awaits further study.

Since the genetic evidence in human and mouse suggests no locus other than the IC carries the germline imprint, it is likely the CTCF sites identified here are bound in response to the IC on the paternal allele. One of the ways in which this can be controlled is through differential DNA methylation (Figure 6-1). Positive influence of the IC on the paternal allele may exclude DNA methylation of CTCF sites and allow binding. Alternatively, heterochromatinization of the maternal allele may recruit DNA methylating activities to these sites. The binding of the flanking CTCF sites may lead to sequestration of the domain into a permissive nuclear address where it could be accessible to modifying enzymes, marking the allele with euchromatic histone modifications. These modifications and differential DNA methylation at the CTCF binding sites may contribute to the somatic memory of the paternal allele throughout development. When the paternal allele arrives in a tissue that carries factors necessary in the transcriptional activation of one of the transposed imprinted genes, they are able to bind the paternal allele at the permissive nuclear address, and affect further accumulation of active histone modifications, leading to opening of the chromatin structure and gene activation. This model fills a void in the understanding of the PWS/AS imprinted domain. As described in the Introduction, the antisense model of imprinting satisfies the observations for many of the imprinted genes such as *SNURF-SNRPN*, snoRNAs and *UBE3A* (Runte et al. 2001). The allelic binding of CTCF at the transposed imprinted genes complements the antisense model with a mechanism to address imprinting of these genes where the antisense mechanism is insufficient.

Figure 6-1. Model of epigenetic relationship between imprinted and tissue-specific regulation.

Schematic of a model of the relationship between chromatin accessibility, histone modifications and CTCF binding to the transposed imprinted genes. An initial signal from the IC in *cis* determines differential epigenetic states on the two alleles, including allelic DNA methylation (lollipop) at CTCF sites (dark green circles) flanking transposed imprinted genes (pentagonal block arrow). This then leads to allelic binding of CTCF (dark blue half donuts), and association of the paternal allele to active regions of the nucleus (green haze), with access to factors that can reinforce a differential epigenetic state. The other allele would lack CTCF binding and be associated with a different set of factors (grey octagons). Tissues expressing one of the transposed imprinted genes would contain tissue-specific factors (orange triangles) that will have access to the paternal allele, and further add active epigenetic marks such as histone modifications and lead to transcription of the paternal allele in that tissue (green pentagonal block arrow).


+ve
Paternal
IC
Dnmts
Maternal
IC
CTCF
CTCF
expressing
Differentiation
non-expressing
+ve
Active nuclear compartment
Active nuclear compartment

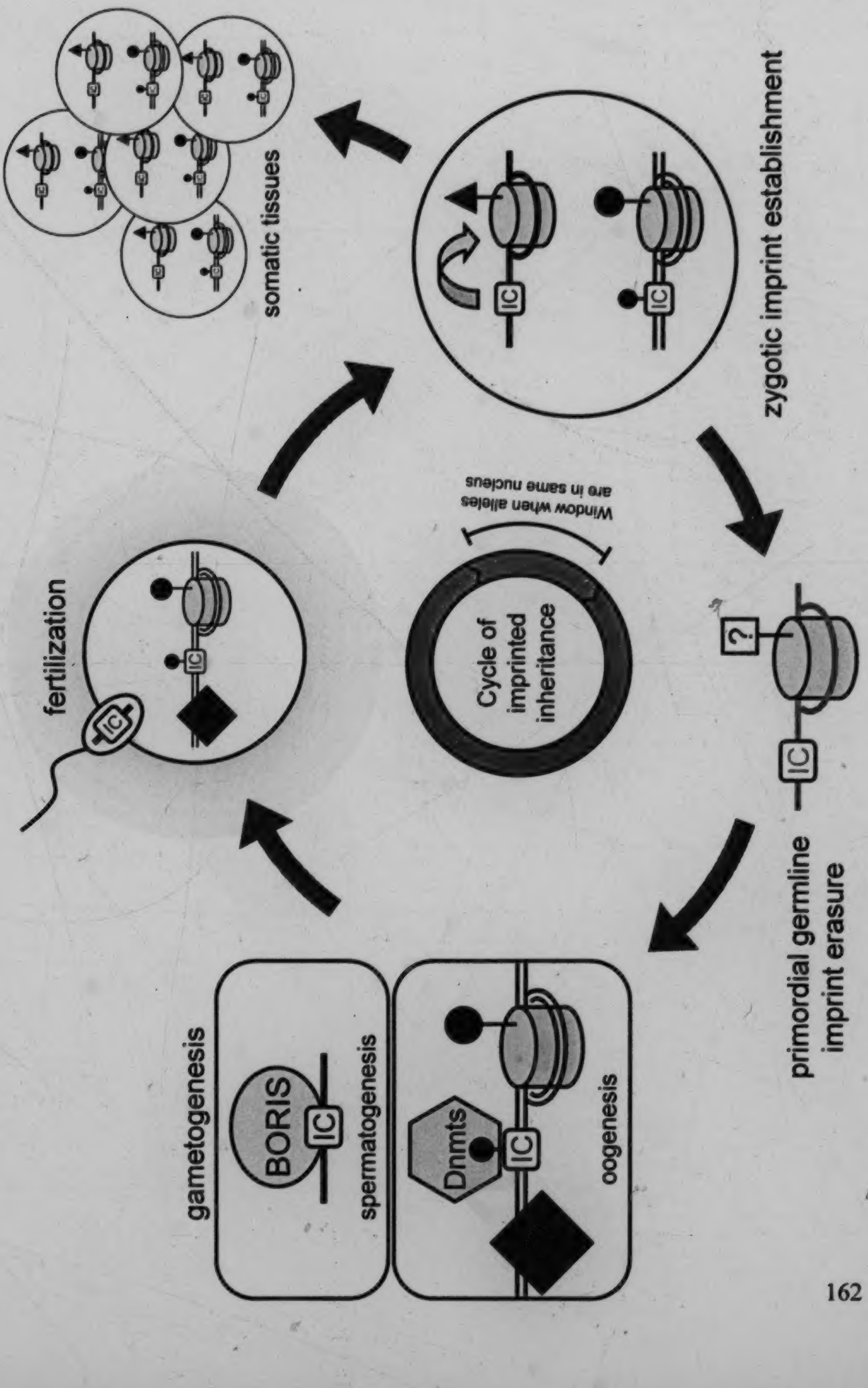
Early events of the transposed imprinted genes

As with all new discoveries in science, answers spawn more questions, and the work presented here will lead to future efforts in filling in details of this model. The importance of determining the order and hierarchy of regulatory events is evident from this work. The experiments described here address the state of imprinting of the transposed imprinted genes in somatic tissues. However, the allelic differences observed were likely not set up *de novo*, but epigenetically inherited. There are two distinct questions here: 1) the question of the order of events early in the setting up and interpretation of the original imprint, and 2) the setup and propagation of the imprint that survives cell divisions in the soma. As was discussed in the Introduction, the order is not even clear at the IC, nor is a complete picture available for any imprinted domain.

Allelic imprints of the previous generation are erased during germline formation through an active global demethylation event (Figure 6-2) (Hajkova et al. 2002). This complete erasure allows each generation to mark imprinted genes according to gender. During gametogenesis, the exclusive expression of the CTCF-related protein, BORIS in the male germline has been postulated to be important in preventing methylation of imprinted ICs that would otherwise be methylated in female gametogenesis (Loukinov et al. 2002). The female germline however, without the need to endure the dramatic genome-wide repackaging required by spermatogenesis, may use DNA methylation, histone modifications or binding of other protein factors to dictate imprinted states. The methylation state of the PWS/AS IC is likely to fit into these schemes and survive the subsequent zygotic demethylation with the maternal IC methylated and paternal IC unmethylated. While DNA methylation for ICs is known to be allelic at this stage, there may be other epigenetic marks present, such as histone modifications. CTCF, the relatively new player on the block, may be involved in the early events in X-inactivation (Pugacheva et al. 2005), but its significance at the PWS/AS IC has not been determined. In general, the hierarchy of control has plagued the field of gene regulation. Instead of a linear hierarchy, there may be a network of overlapping epigenetic signals reinforcing and fine tuning the regulation of a gene. This view may be a compromise but still does not suggest a candidate initiating signal driving allelic expression of the transposed imprinted genes. Detailed description of the correlation between other imprinted genes and between different species in terms of the language of epigenetic modifications will also help our understanding of the importance of different marks over others. Undoubtedly, whole genome approaches will be instrumental in this debate (Bernstein et al. 2005).

Figure 6-2. Cycle of imprinted inheritance of the IC.

Establishment of the imprinted state as observed in somatic tissues (top right) begins in development of the gametes in the previous generation (left). Blue and red lines represent DNA of the paternal and maternal allele, respectively. Differential DNA methylation patterns at ICs are the result of events specific to spermatogenesis and oogenesis. In spermatogenesis, testis-specific factors such as BORIS may play a role in determining DNA methylation states, whereas in the ovary, factors other than DNA methylation may also be employed such as histone modifications or binding of unknown proteins (green diamond). After fertilization and the mixture of the two parental contributions, these epigenetic states must then be elaborated into other epigenetic signals that aid in transmission and transcriptional control of the two alleles. Cells of the soma in the new generation must re-enter this cycle of imprint inheritance to maintain correct gender-specific marks for the next generation.


gametogenesis
BORIS
IC
spermatogenesis
Dnmts
IC
oogenesis
fertilization
IC
IC
Cycle of imprinted inheritance
Window when alleles are in same nucleus
zygotic imprint establishment
IC
IC
somatic tissues
primordial germline imprint erasure
IC
?

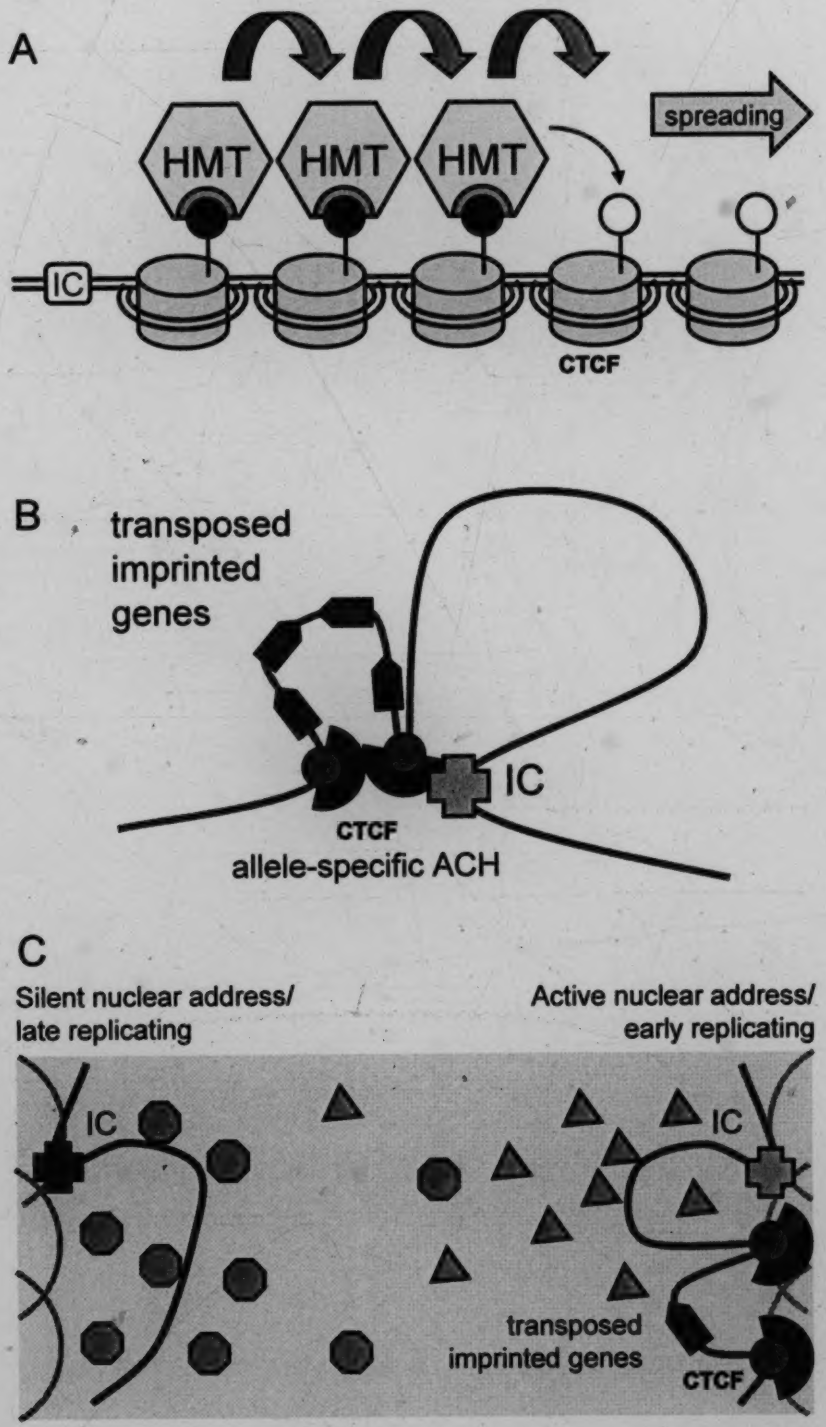
The allelic identity at the IC must also traverse the genomic distance necessary to reach the imprinted genes in the PWS/AS region. Evidence in PWS patients with somatic deletions of the IC indicates that this is an ongoing process that requires influence of the IC throughout development, instead of an isolated early post-fertilization event (Bielinska et al. 2000). While our studies in somatic tissues only indirectly addressed the initial events determining allelic identity, somatic tissues are ideal for studying the constant requirement of the transposed imprinted genes for an LCR-like function for the IC in imprinting.

Talking to the transposed imprinted genes

Our question of hierarchy is intimately linked to another outstanding question of wide interest in imprinting and gene regulation, and that is how regulatory signals traverse significant genomic distances. Mechanisms involved in both forms of regulation may be one and the same. One of the simplest and oldest models of long range control is physical spreading of epigenetic state. The patterns of DNA methylation and histone methylation that are thought to function in transcriptional regulation may also play the role of long distance communication. This can be achieved through reinforcement of the epigenetic network as described above and in the Introduction (Figure 1-11). For example, H3K9me3 is able to bind the heterochromatic protein HP1, which in turn is able to recruit the H3K9 HMT SUV39H1, which then reinforces the positive feedback loop by methylating H3K9 (Figure 6-3 A) (Grewal and Moazed 2003). This loop is not the only mechanism however, as H3K9me can cause histone deacetylation and silencing independently of HP1 (Stewart et al. 2005).

Figure 6-3. Models of distant epigenetic influence.

Depicted here are several non-mutually exclusive models of distant control by an LCR or IC element. (A) Model of linear spreading of epigenetic modifications, with propagation of H3K9me schematically shown. The epigenetic loop of H3K9me binding of the HMT SUV39H1, subsequent methylation of H3K9 and further recruitment of HMT can feedback and amplify the spreading of the heterochromatic state on the maternal allele from the IC. (B) Model of direct communication between the IC and the transposed imprinted gene cluster (green and red pentagonal arrows) by intralocus association. A protein of unknown identity (orange cross) may bind the IC and form an active chromatin hub (ACH) with CTCF (blue half donuts) bound at a distance. (C) Model of physical or temporal separation of alleles by action of the IC. The IC may cause allelic sequestration of the alleles into different nuclear addresses or differential replication timing. Separation would result in differential access to gradients of activating complexes (green triangles), and silencing complexes (grey octagons).


A
HMT
HMT
HMT
spreading
IC
CTCF
B
transposed imprinted genes
IC
CTCF
allele-specific ACH
C
Silent nuclear address/ late replicating
Active nuclear address/ early replicating
IC
IC
transposed imprinted genes
CTCF

Looping mechanisms such as described for *Igf2/H19* are one of the most common models for long distance control (Bulger and Groudine 1999). Although our data do not support CTCF binding at the IC, the data do not preclude intralocus association between the IC and the CTCF sites flanking the transposed imprinted genes. In fact, it is clear that factors other than CTCF must be able to participate in the loop formation at *Igf2/H19* since CTCF only binds the unmethylated DMR, but the methylated DMR, and other sequences found to associate with the DMR in *cis* do not bind CTCF (Murrell et al. 2004). The identity of these DNA binding proteins are unknown, but it is possible that similar factors can bind the PWS IC in a similar way and mediate association with the CTCF sites flanking the transposed imprinted genes, thus forming an allele-specific active chromatin hub (Figure 6-3 C). It is surprising that no reports detailing the intralocus association of the PWS/AS or any other imprinted domain has been published thus far. However, lack of supporting evidence does not equate to evidence to the contrary, and may simply be due to the technical difficulties involved in assaying a region as large as the PWS/AS domain with present techniques. Intralocus association would certainly explain why the IC is necessary for continued imprinting of the region in somatic tissues. Further characterization of these associations and factors involved will be critical in understanding imprinting of the PWS/AS cluster and how this relates to imprinting in general.

Another perspective in viewing the question of how the IC imparts its allelic signal is to look at the physical separation of the alleles. Initial erasure and establishment of parental imprints occurs in gametogenesis when parental alleles are isolated. Even after fertilization, the parental genomes remain separate in a stage when germline imprints may be translated to other allelic epigenetic marks. After these stages, it may be a challenge to physically separate the alleles to facilitate differential regulation. A model of sequestration of alleles based on CTCF binding has already been presented here. If the IC functions in a similar way, by allelically determining structural nuclear association, as suggested by others (Greally et al. 1999; Kagotani et al. 2002), this would be another mechanism consistent with a somatic requirement of the IC. An alternative involves temporal separation. Replication timing has been associated with imprinted domains but has not been well characterized in terms of its function in imprinting (Bickmore and Carothers 1995). This represents a unique mechanism of separating alleles. Perhaps histone modifying and chromatin remodeling activities vary during the cell cycle and alleles replicating at different times may gain or only be able to maintain certain states. While the replication timing for the IC has been studied, replication of the transposed imprinted genes could be under a different origin of replication and its timing may be dependent on other factors. Physical or temporal separation of alleles would also represent a solution to long range communication between the IC and the transposed imprinted genes; by simple indirect influence. Communication may not be necessary if the IC can direct the two alleles to completely different nuclear addresses and limit the potential chromatin modifying and transcriptional machinery available. All of these questions will fuel the unfolding of this biological puzzle for years to come.

References

Adamson ED, Yu J, Mustelin T (2005) Co-factors p300 and CBP catch egr1 in their network. Prostate 63:407-410

Ahmad K, Henikoff S (2002) The histone variant h3.3 marks active chromatin by replication- independent nucleosome assembly. Mol Cell 9:1191-1200

Ahn SH, Cheung WL, Hsu JY, Diaz RL, Smith MM, Allis CD (2005) Sterile 20 Kinase Phosphorylates Histone H2B at Serine 10 during Hydrogen Peroxide-Induced Apoptosis in S. cerevisiae. Cell 120:25-36

Arney KL (2003) H19 and Igf2 - enhancing the confusion? Trends Genet 19:17-23

Arranz V, Dreuillet C, Crisanti P, Tillit J, Kress M, Ernoult-Lange M (2001) The zinc finger transcription factor, MOK2, negatively modulates expression of the interphotoreceptor retinoid-binding protein gene, IRBP. J Biol Chem 276:11963-11969

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds) (1993) Current Protocols in Molecular Biology. Vol 1. Current Protocols, New York, N.Y.

Awad TA, Bigler J, Ulmer JE, Hu YJ, Moore JM, Lutz M, Neiman PE, Collins SJ, Renkawitz R, Lobanenkov VV, Filippova GN (1999) Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel target DNA sequence. J Biol Chem 274:27092-27098

Balmer D, Arredondo J, Samaco RC, LaSalle JM (2002) MECP2 mutations in Rett syndrome adversely affect lymphocyte growth, but do not affect imprinted gene expression in blood or brain. Hum Genet 110:545-552

Baniahmad A, Steiner C, Kohne AC, Renkawitz R (1990) Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. Cell 61:505-514

Bannister AJ, Schneider R, Kouzarides T (2002) Histone methylation: dynamic or static? Cell 109:801-806

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410:120-124

Bartee L, Malagnac F, Bender J (2001) Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev 15:1753-1758

Bartolomei MS, Zemel S, Tilghman SM (1991) Parental imprinting of the mouse H19 gene. Nature 351:153-155

Beard C, Li E, Jaenisch R (1995) Loss of methylation activates Xist in somatic but not in embryonic cells. Genes Dev 9:2325-2334

Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405:482-485

Bell AC, West AG, Felsenfeld G (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98:387-396

Berger SL (2001) Molecular biology. The histone modification circus. Science 292:64-65

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, 3rd, Gingeras TR, Schreiber SL, Lander ES (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120:169-181

Bestor TH (2000) The DNA methyltransferases of mammals. Hum Mol Genet 9:2395-2402

Bhattacharya SK, Ramchandani S, Cervoni N, Szyf M (1999) A mammalian protein with specific demethylase activity for mCpG DNA. Nature 397:579-583

Bickmore WA, Carothers AD (1995) Factors affecting the timing and imprinting of replication on a mammalian chromosome. Journal of Cell Science 108:2801-2809

Bielinska B, Blaydes SM, Buiting K, Yang T, Krajewska-Walasek M, Horsthemke B, Brannan CI (2000) De novo deletions of SNRPN exon 1 in early human and mouse embryos result in a paternal to maternal imprint switch. Nat Genet 25:74-78

Bienvenu T, Carrie A, de Roux N, Vinet MC, Jonveaux P, Couvert P, Villard L, Arzimanoglou A, Beldjord C, Fontes M, Tardieu M, Chelly J (2000) MECP2 mutations account for most cases of typical forms of Rett syndrome. Hum Mol Genet 9:1377-1384

Biniszkiewicz D, Gribnau J, Ramsahoye B, Gaudet F, Eggan K, Humpherys D, Mastrangelo MA, Jun Z, Walter J, Jaenisch R (2002) Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. Mol Cell Biol 22:2124-2135

Bird A (1997) Does DNA methylation control transposition of selfish elements in the germline? Trends Genet 13:469-472

Bird AP, Wolffe AP (1999) Methylation-induced repression--belts, braces, and chromatin. Cell 99:451-454

Boggs BA, Cheung P, Heard E, Spector DL, Chinault AC, Allis CD (2002) Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nat Genet 30:73-76

Bonne-Andrea C, Wong ML, Alberts BM (1990) In vitro replication through nucleosomes without histone displacement. Nature 343:719-726

Borson ND, Lacy MQ, Wettstein PJ (2002) Altered mRNA expression of Pax5 and Blimp-1 in B cells in multiple myeloma. Blood 100:4629-4639

Brannan CI, Bartolomei MS (1999) Mechanisms of genomic imprinting. Curr Opin Genet Dev 9:164-170

Bressler J, Tsai TF, Wu MY, Tsai SF, Ramirez MA, Armstrong D, Beaudet AL (2001) The SNRPN promoter is not required for genomic imprinting of the Prader-Willi/Angelman domain in mice. Nat Genet 28:232-240

Buiting K, Dittrich B, Endele S, Horsthemke B (1997) Identification of novel exons 3' to the human SNRPN gene. Genomics 40:132-137

Buiting K, Gross S, Lich C, Gillessen-Kaesbach G, el-Maarri O, Horsthemke B (2003) Epimutations in Prader-Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect. Am J Hum Genet 72:571-577

Buiting K, Saitoh S, Gross S, Dittrich B, Schwartz S, Nicholls RD, Horsthemke B (1995) Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. Nat Genet 9:395-400

Bulger M, Groudine M (1999) Looping versus linking: toward a model for long-distance gene activation. Genes Dev 13:2465-2477

Burger J, Horn D, Tonnies H, Neitzel H, Reis A (2002) Familial interstitial 570 kbp deletion of the UBE3A gene region causing Angelman syndrome but not Prader-Willi syndrome. Am J Med Genet 111:233-237

Busslinger M (2004) Transcriptional control of early B cell development. Annu Rev Immunol 22:55-79

Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P (2002) Long-range chromatin regulatory interactions in vivo. Nat Genet 11:11

Caspary T, Cleary MA, Baker CC, Guan XJ, Tilghman SM (1998) Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. Mol Cell Biol 18:3466-3474

Cattanach BM, Barr JA, Beechey CV, Martin J, Noebels J, Jones J (1997) A candidate model for Angelman syndrome in the mouse. Mamm Genome 8:472-478

Cattanach BM, Barr JA, Evans EP, Burtenshaw M, Beechey CV, Leff SE, Brannan CI, Copeland NG, Jenkins NA, Jones J (1992) A candidate mouse model for Prader-Willi syndrome which shows an absence of Snrpn expression. Nat Genet 2:270-274

Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc Natl Acad Sci U S A 97:14311-14316

Chadwick BP, Willard HF (2001) A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. J Cell Biol 152:375-384

Chadwick BP, Willard HF (2002) Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome. J Cell Biol 157:1113-1123

Chai JH, Locke DP, Ohta T, Greally JM, Nicholls RD (2001) Retrotransposed genes such as Frat3 in the mouse Chromosome 7C Prader- Willi syndrome region acquire the imprinted status of their insertion site. Mamm Genome 12:813-821

Chakravarti D, LaMorte VJ, Nelson MC, Nakajima T, Schulman IG, Juguilon H, Montminy M, Evans RM (1996) Role of CBP/P300 in nuclear receptor signalling. Nature 383:99-103

Chamberlain SJ, Brannan CI (2001) The Prader-Willi syndrome imprinting center activates the paternally expressed murine Ube3a antisense transcript but represses paternal Ube3a. Genomics 73:316-322

Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT (2002) CTCF, a candidate trans-acting factor for X-inactivation choice. Science 295:345-347

Chen D, Ma H, Hong H, Koh SS, Huang SM, Schurter BT, Aswad DW, Stallcup MR (1999) Regulation of transcription by a protein methyltransferase. Science 284:2174-2177

Chen RZ, Akbarian S, Tudor M, Jaenisch R (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat Genet 27:327-331

Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM, Allis CD (2000) Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell 5:905-915

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156-159

Christophe D (2004) The control of thyroid-specific gene expression: what exactly have we learned as yet? Mol Cell Endocrinol 223:1-4

Cohen-Armon M, Visochek L, Katzoff A, Levitan D, Susswein AJ, Klein R, Valbrun M, Schwartz JH (2004) Long-term memory requires polyADP-ribosylation. Science 304:1820-1822

Craig JM (2005) Heterochromatin-many flavours, common themes. Bioessays 27:17-28

Czerny T, Busslinger M (1995) DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). Mol Cell Biol 15:2858-2871

da Rocha ST, Ferguson-Smith AC (2004) Genomic imprinting. Curr Biol 14:R646-649

Davie JR, Murphy LC (1994) Inhibition of transcription selectively reduces the level of ubiquitinated histone H2B in chromatin. Biochem Biophys Res Commun 203:344-350

de la Cruz X, Lois S, Sanchez-Molina S, Martinez-Balbas MA (2005) Do protein motifs read the histone code? Bioessays 27:164-175

de los Santos T, Schweizer J, Rees CA, Francke U (2000) Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from PWCR1, a novel imprinted gene in the Prader- Willi deletion region, which Is highly expressed in brain. Am J Hum Genet 67:1067-1082

DeChiara TM, Robertson EJ, Efstratiadis A (1991) Parental imprinting of the mouse insulin-like growth factor II gene. Cell 64:849-859

Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295:1306-1311

Delaval K, Feil R (2004) Epigenetic regulation of mammalian genomic imprinting. Curr Opin Genet Dev 14:188-195

Dion MF, Altschuler SJ, Wu LF, Rando OJ (2005) From the Cover: Genomic characterization reveals a simple histone H4 acetylation code. Proc Natl Acad Sci U S A 102:5501-5506

Dittmer J (2003) The biology of the Ets1 proto-oncogene. Mol Cancer 2:29

Dittrich B, Buiting K, Korn B, Rickard S, Buxton J, Saitoh S, Nicholls RD, Poustka A, Winterpacht A, Zabel B, Horsthemke B (1996) Imprint switching on human chromosome 15 may involve alternative transcripts of the SNRPN gene. Nat Genet 14:163-170

Dreuillet C, Tillit J, Kress M, Ernoult-Lange M (2002) In vivo and in vitro interaction between human transcription factor MOK2 and nuclear lamin A/C. Nucleic Acids Res 30:4634-4642

Eccles MR, He S, Legge M, Kumar R, Fox J, Zhou C, French M, Tsai RW (2002) PAX genes in development and disease: the role of PAX2 in urogenital tract development. Int J Dev Biol 46:535-544

Eden S, Cedar H (1994) Role of DNA methylation in the regulation of transcription. Curr Opin Genet Dev 4:255-259

El Kharroubi A, Piras G, Stewart CL (2001) DNA demethylation reactivates a subset of imprinted genes in uniparental mouse embryonic fibroblasts. J Biol Chem 276:8674-8680

El-Maarri O, Buiting K, Peery EG, Kroisel PM, Balaban B, Wagner K, Urman B, Heyd J, Lich C, Brannan CI, Walter J, Horsthemke B (2001) Maternal methylation imprints on human chromosome 15 are established during or after fertilization. Nat Genet 27:341-344

El-Osta A, Wolffe AP (2000) DNA methylation and histone deacetylation in the control of gene expression: basic biochemistry to human development and disease. Gene Expr 9:63-75

Ernoult-Lange M, Kress M, Hamer D (1990) A gene that encodes a protein consisting solely of zinc finger domains is preferentially expressed in transformed mouse cells. Mol Cell Biol 10:418-421

Farber C, Dittrich B, Buiting K, Horsthemke B (1999) The chromosome 15 imprinting centre (IC) region has undergone multiple duplication events and contains an upstream exon of SNRPN that is deleted in all Angelman syndrome patients with an IC microdeletion. Hum Mol Genet 8:337-343

Felsenfeld G, Groudine M (2003) Controlling the double helix. Nature 421:448-453

Ferguson-Smith AC, Surani MA (2001) Imprinting and the epigenetic asymmetry between parental genomes. Science 293:1086-1089

Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV (1996) An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol 16:2802-2813

Filippova GN, Qi CF, Ulmer JE, Moore JM, Ward MD, Hu YJ, Loukinov DI, Pugacheva EM, Klenova EM, Grundy PE, Feinberg AP, Cleton-Jansen AM, Moerland EW, Cornelisse CJ, Suzuki H, Komiya A, Lindblom A, Dorion-Bonnet F, Neiman PE, Morse HC, 3rd, Collins SJ, Lobanenkov VV (2002) Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. Cancer Res 62:48-52

Fischle W, Wang Y, Allis CD (2003a) Binary switches and modification cassettes in histone biology and beyond. Nature 425:475-479

Fischle W, Wang Y, Allis CD (2003b) Histone and chromatin cross-talk. Curr Opin Cell Biol 15:172-183

Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S (2003c) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev 17:1870-1881

Fitzpatrick GV, Soloway PD, Higgins MJ (2002) Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat Genet 9:9

Forsberg EC, Downs KM, Christensen HM, Im H, Nuzzi PA, Bresnick EH (2000) Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. Proc Natl Acad Sci U S A 97:14494-14499

Fournier C, Goto Y, Ballestar E, Delaval K, Hever AM, Esteller M, Feil R (2002) Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. EMBO J 21:6560-6570

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem 278:4035-4040

Fulmer-Smentek SB, Francke U (2001) Association of acetylated histones with paternally expressed genes in the Prader--Willi deletion region. Hum Mol Genet 10:645-652

Gabriel JM, Merchant M, Ohta T, Ji Y, Caldwell RG, Ramsey MJ, Tucker JD, Longnecker R, Nicholls RD (1999) A transgene insertion creating a heritable chromosome deletion mouse model of Prader-Willi and angelman syndromes. Proc Natl Acad Sci U S A 96:9258-9263

Gerard M, Hernandez L, Wevrick R, Stewart CL (1999) Disruption of the mouse necdin gene results in early post-natal lethality. Nat Genet 23:199-202

Gerasimova TI, Byrd K, Corces VG (2000) A chromatin insulator determines the nuclear localization of DNA. Mol Cell 6:1025-1035

Girdwood D, Bumpass D, Vaughan OA, Thain A, Anderson LA, Snowden AW, Garcia-Wilson E, Perkins ND, Hay RT (2003) P300 transcriptional repression is mediated by SUMO modification. Mol Cell 11:1043-1054

Glenn CC, Saitoh S, Jong MT, Filbrandt MM, Surti U, Driscoll DJ, Nicholls RD (1996) Gene structure, DNA methylation, and imprinted expression of the human SNRPN gene. Am J Hum Genet 58:335-346

Goren A, Cedar H (2003) Replicating by the clock. Nat Rev Mol Cell Biol 4:25-32

Grandjean V, O'Neill L, Sado T, Turner B, Ferguson-Smith A (2001) Relationship between DNA methylation, histone H4 acetylation and gene expression in the mouse imprinted Igf2-H19 domain. FEBS Lett 488:165-169

Grant PA (2001) A tale of histone modifications. Genome Biol 2:REVIEWS0003

Gray TA, Saitoh S, Nicholls RD (1999) An imprinted, mammalian bicistronic transcript encodes two independent proteins. Proc Natl Acad Sci U S A 96:5616-5621

Greally JM, Gray TA, Gabriel JM, Song L, Zemel S, Nicholls RD (1999) Conserved characteristics of heterochromatin-forming DNA at the 15q11- q13 imprinting center. Proc Natl Acad Sci U S A 96:14430-14435

Grewal SI, Moazed D (2003) Heterochromatin and epigenetic control of gene expression. Science 301:798-802

Grunstein M (1997) Histone acetylation in chromatin structure and transcription. Nature 389:349-352

Guy J, Hendrich B, Holmes M, Martin JE, Bird A (2001) A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet 27:322-326

Haig D (2004) Genomic imprinting and kinship: how good is the evidence? Annu Rev Genet 38:553-585

Haig D, Graham C (1991) Genomic imprinting and the strange case of the insulin-like growth factor II receptor. Cell 64:1045-1046

Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani M (2002) Epigenetic reprogramming in mouse primordial germ cells. Mech Dev 117:15-23

Hanel ML (2003) Epigenetic modifications associated with the maternal and paternal alleles of necdin, an imprinted gene in the Prader-Willi Syndrome region. Doctor of Philosophy, University of Alberta, Edmonton

Hanel ML, Lau JC, Paradis I, Drouin R, Wevrick R (2005) Chromatin modification of the human imprinted NDN (necdin) gene detected by in vivo footprinting. J Cell Biochem 94:1046-1057

Hanel ML, Wevrick R (2001) Establishment and maintenance of DNA methylation patterns in mouse Ndn: implications for maintenance of imprinting in target genes of the imprinting center. Mol Cell Biol 21:2384-2392

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A 96:14412-14417

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405:486-489

Hark AT, Tilghman SM (1998) Chromatin conformation of the H19 epigenetic mark. Hum Mol Genet 7:1979-1985

Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL (2001) Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell 107:727-738

Hebbes TR, Clayton AL, Thorne AW, Crane-Robinson C (1994) Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. EMBO J 13:1823-1830

Hebbes TR, Thorne AW, Crane-Robinson C (1988) A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J 7:1395-1402

Hendrich B, Tweedie S (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet 19:269-277

Henikoff S, Furuyama T, Ahmad K (2004) Histone variants, nucleosome assembly and epigenetic inheritance. Trends Genet 20:320-326

Holm VA, Cassidy SB, Butler MG, Hanchett JM, Greenswag LR, Whitman BY, Greenberg F (1993) Prader-Willi syndrome: consensus diagnostic criteria. Pediatrics 91:398-402

Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T (2005) Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Genet 37:31-40

Horike S, Mitsuya K, Meguro M, Kotobuki N, Kashiwagi A, Notsu T, Schulz TC, Shirayoshi Y, Oshimura M (2000) Targeted disruption of the human LIT1 locus defines a putative imprinting control element playing an essential role in Beckwith-Wiedemann syndrome. Hum Mol Genet 9:2075-2083

Horsthemke B (1997) Structure and function of the human chromosome 15 imprinting center. J Cell Physiol 173:237-241

Hsu JY, Sun ZW, Li X, Reuben M, Tatchell K, Bishop DK, Grushcow JM, Brame CJ, Caldwell JA, Hunt DF, Lin R, Smith MM, Allis CD (2000) Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. Cell 102:279-291

Hu JF, Oruganti H, Vu TH, Hoffman AR (1998) The role of histone acetylation in the allelic expression of the imprinted human insulin-like growth factor II gene. Biochem Biophys Res Commun 251:403-408

Iizuka M, Smith MM (2003) Functional consequences of histone modifications. Curr Opin Genet Dev 13:154-160

Ishihara K, Sasaki H (2002) An evolutionarily conserved putative insulator element near the 3' boundary of the imprinted Igf2/H19 domain. Hum Mol Genet 11:1627-1636

Jaenisch R (1997) DNA methylation and imprinting: why bother? Trends Genet 13:323-329

Jason LJ, Moore SC, Lewis JD, Lindsey G, Ausio J (2002) Histone ubiquitination: a tagging tail unfolds? Bioessays 24:166-174

Jay P, Rougeulle C, Massacrier A, Moncla A, Mattei MG, Malzac P, Roeckel N, Taviaux S, Lefranc JL, Cau P, Berta P, Lalande M, Muscatelli F (1997) The human necdin gene, NDN, is maternally imprinted and located in the Prader-Willi syndrome chromosomal region. Nat Genet 17:357-361

Jenuwein T, Allis CD (2001) Translating the histone code. Science 293:1074-1080

Johnson DR (1974) Hairpin-tail: a case of post-reductional gene action in the mouse egg. Genetics 76:795-805

Johnson K, Pflugh DL, Yu D, Hesslein DG, Lin KI, Bothwell AL, Thomas-Tikhonenko A, Schatz DG, Calame K (2004) B cell-specific loss of histone 3 lysine 9 methylation in the V(H) locus depends on Pax5. Nat Immunol 5:853-861

Johnson L, Cao X, Jacobsen S (2002) Interplay between Two Epigenetic Marks. DNA Methylation and Histone H3 Lysine 9 Methylation. Curr Biol 12:1360-1367

Jones PA, Takai D (2001) The role of DNA methylation in mammalian epigenetics. Science 293:1068-1070

Jong MT, Gray TA, Ji Y, Glenn CC, Saitoh S, Driscoll DJ, Nicholls RD (1999) A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region. Hum Mol Genet 8:783-793

Jonkers J, van Amerongen R, van der Valk M, Robanus-Maandag E, Molenaar M, Destree O, Berns A (1999) In vivo analysis of Frat1 deficiency suggests compensatory activity of Frat3. Mech Dev 88:183-194

Kagotani K, Nabeshima H, Kohda A, Nakao M, Taguchi H, Okumura K (2002) Visualization of transcription-dependent association of imprinted genes with the nuclear matrix. Exp Cell Res 274:189-196

Kantor B, Kaufman Y, Makedonski K, Razin A, Shemer R (2004a) Establishing the epigenetic status of the Prader-Willi/Angelman imprinting center in the gametes and embryo. Hum Mol Genet 13:2767-2779

Kantor B, Makedonski K, Green-Finberg Y, Shemer R, Razin A (2004b) Control elements within the PWS/AS imprinting box and their function in the imprinting process. Hum Mol Genet 13:751-762

Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ (2003) The UCSC Genome Browser Database. Nucleic Acids Res 31:51-54

Kato Y, Sasaki H (2005) Imprinting and looping: epigenetic marks control interactions between regulatory elements. Bioessays 27:1-4

Keohane AM, Lavender JS, O'Neill LP, Turner BM (1998) Histone acetylation and X inactivation. Dev Genet 22:65-73

Kimura H, Shiota K (2003) Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. J Biol Chem 278:4806-4812

Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, Farnham PJ (2004) Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev 18:1592-1605

Klenova EM, Morse HC, 3rd, Ohlsson R, Lobanenkov VV (2002) The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. Semin Cancer Biol 12:399-414

Knoll JH, Cheng SD, Lalande M (1994) Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region. Nat Genet 6:41-46

Knoll JH, Nicholls RD, Magenis RE, Graham JM, Jr., Lalande M, Latt SA (1989) Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. Am J Med Genet 32:285-290

Kobayashi A, Yamagiwa H, Hoshino H, Muto A, Sato K, Morita M, Hayashi N, Yamamoto M, Igarashi K (2000) A combinatorial code for gene expression generated by transcription factor Bach2 and MAZR (MAZ-related factor) through the BTB/POZ domain. Mol Cell Biol 20:1733-1746

Kobayashi S, Kohda T, Ichikawa H, Ogura A, Ohki M, Kaneko-Ishino T, Ishino F (2002) Paternal expression of a novel imprinted gene, Peg12/Frat3, in the mouse 7C region homologous to the Prader-Willi syndrome region. Biochem Biophys Res Commun 290:403-408

Koritschoner NP, Bocco JL, Panzetta-Dutari GM, Dumur CI, Flury A, Patrito LC (1997) A novel human zinc finger protein that interacts with the core promoter element of a TATA box-less gene. J Biol Chem 272:9573-9580

Krude T (1999) Chromatin assembly during DNA replication in somatic cells. Eur J Biochem 263:1-5

Kuo MH, Allis CD (1999) In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. Methods 19:425-433

Kuo MH, Brownell JE, Sobel RE, Ranalli TA, Cook RG, Edmondson DG, Roth SY, Allis CD (1996) Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature 383:269-272

Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410:116-120

Lacoste N, Utley RT, Hunter JM, Poirier GG, Cote J (2002) Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. J Biol Chem 277:30421-30424

Landers M, Bancescu DL, Le Meur E, Rougeulle C, Glatt-Deeley H, Brannan C, Muscatelli F, Lalande M (2004) Regulation of the large (approximately 1000 kb) imprinted murine Ube3a antisense transcript by alternative exons upstream of Snurf/Snrpn. Nucleic Acids Res 32:3480-3492

Lau JC, Hanel ML, Wevrick R (2004) Tissue-specific and imprinted epigenetic modifications of the human NDN gene. Nucleic Acids Res 32:3376-3382

Lee JT (2003) Molecular Links between X-Inactivation and Autosomal Imprinting: X-Inactivation as a Driving Force for the Evolution of Imprinting? Curr Biol 13:R242-254

Lee S, Kozlov S, Hernandez L, Chamberlain SJ, Brannan CI, Stewart CL, Wevrick R (2000) Expression and imprinting of MAGEL2 suggest a role in Prader-willi syndrome and the homologous murine imprinting phenotype. Hum Mol Genet 9:1813-1819

Lee S, Walker CL, Karten B, Kuny SL, Tennese AA, O'Neill MA, Wevrick R (2005) Essential role for the Prader-Willi syndrome protein necdin in axonal outgrowth. Hum Mol Genet 14:627-637

Lee S, Wevrick R (2000) Identification of novel imprinted transcripts in the Prader-Willi syndrome and Angelman syndrome deletion region: further evidence for regional imprinting control. Am J Hum Genet 66:848-858

Leighton PA, Ingram RS, Eggenschwiler J, Efstratiadis A, Tilghman SM (1995) Disruption of imprinting caused by deletion of the H19 gene region in mice. Nature 375:34-39

Li E, Beard C, Jaenisch R (1993) Role for DNA methylation in genomic imprinting. Nature 366:362-365

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915-926

Li Q, Harju S, Peterson KR (1999) Locus control regions: coming of age at a decade plus. Trends Genet 15:403-408

Lindroth AM, Shultis D, Jasencakova Z, Fuchs J, Johnson L, Schubert D, Patnaik D, Pradhan S, Goodrich J, Schubert I, Jenuwein T, Khorasanizadeh S, Jacobsen SE (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. EMBO J 23:4146-4155

Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G (2001) Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. Science 293:2453-2455

Lo WS, Trievel RC, Rojas JR, Duggan L, Hsu JY, Allis CD, Marmorstein R, Berger SL (2000) Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. Mol Cell 5:917-926

Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH (1990) A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene 5:1743-1753

Loukinov DI, Pugacheva E, Vatolin S, Pack SD, Moon H, Chernukhin I, Mannan P, Larsson E, Kanduri C, Vostrov AA, Cui H, Niemitz EL, Rasko JE, Docquier FM, Kistler M, Breen JJ, Zhuang Z, Quitschke WW, Renkawitz R, Klenova EM, Feinberg AP, Ohlsson R, Morse HC, 3rd, Lobanenkov VV (2002) BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. Proc Natl Acad Sci U S A 99:6806-6811

Lucchini R, Wellinger RE, Sogo JM (2001) Nucleosome positioning at the replication fork. EMBO J 20:7294-7302

Lutz M, Burke LJ, Barreto G, Goeman F, Greb H, Arnold R, Schultheiss H, Brehm A, Kouzarides T, Lobanenkov V, Renkawitz R (2000) Transcriptional repression by the insulator protein CTCF involves histone deacetylases. Nucleic Acids Res 28:1707-1713

Lyon MF (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190:372-373

MacDonald HR, Wevrick R (1997) The necdin gene is deleted in Prader-Willi syndrome and is imprinted in human and mouse. Hum Mol Genet 6:1873-1878

Magdinier F, Yusufzai TM, Felsenfeld G (2004) Both CTCF-dependent and -independent insulators are found between the mouse T cell receptor alpha and Dad1 genes. J Biol Chem 279:25381-25389

Makedonski K, Abuhatzira L, Kaufman Y, Razin A, Shemer R (2005) MeCP2 deficiency in Rett syndrome causes epigenetic aberrations at the PWS/AS imprinting center that affects UBE3A expression. Hum Mol Genet 14:1049-1058

Mancini-DiNardo D, Steele SJ, Ingram RS, Tilghman SM (2003) A differentially methylated region within the gene Kcnq1 functions as an imprinted promoter and silencer. Hum Mol Genet 12:283-294

Mann JR, Szabo PE, Reed MR, Singer-Sam J (2000) Methylated DNA sequences in genomic imprinting. Crit Rev Eukaryot Gene Expr 10:241-257

Martienssen R (1998) Transposons, DNA methylation and gene control. Trends Genet 14:263-264

Matzke MA, Birchler JA (2005) RNAi-mediated pathways in the nucleus. Nat Rev Genet 6:24-35

Mayer W, Niveleau A, Walter J, Fundele R, Haaf T (2000) Demethylation of the zygotic paternal genome. Nature 403:501-502

McGrath J, Solter D (1984a) Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37:179-183

McGrath J, Solter D (1984b) Maternal Thp lethality in the mouse is a nuclear, not cytoplasmic, defect. Nature 308:550-551

McKittrick E, Gafken PR, Ahmad K, Henikoff S (2004) From The Cover: Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci U S A 101:1525-1530

Meehan RR (2003) DNA methylation in animal development. Semin Cell Dev Biol 14:53-65

Meguro M, Kashiwagi A, Mitsuya K, Nakao M, Kondo I, Saitoh S, Oshimura M (2001) A novel maternally expressed gene, ATP10C, encodes a putative aminophospholipid translocase associated with Angelman syndrome. Nat Genet 28:19-20

Mermoud JE, Popova B, Peters AH, Jenuwein T, Brockdorff N (2002) Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation. Curr Biol 12:247-251

Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, Keightley PD, Bishop SM, Clarke AR, Bird A (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science 297:403-405

Morgan HD, Santos F, Green K, Dean W, Reik W (2005) Epigenetic reprogramming in mammals. Hum Mol Genet 14 Suppl 1:R47-58

Morison IM, Reeve AE (1998) A catalogue of imprinted genes and parent-of-origin effects in humans and animals. Hum Mol Genet 7:1599-1609

Morrison AJ, Highland J, Krogan NJ, Arbel-Eden A, Greenblatt JF, Haber JE, Shen X (2004) INO80 and gamma-H2AX Interaction Links ATP-Dependent Chromatin Remodeling to DNA Damage Repair. Cell 119:767-775

Mowery-Rushton PA, Driscoll DJ, Nicholls RD, Locker J, Surti U (1996) DNA methylation patterns in human tissues of uniparental origin using a zinc-finger gene (ZNF127) from the Angelman/Prader-Willi region. Am J Med Genet 61:140-146

Mukhopadhyay R, Yu W, Whitehead J, Xu J, Lezcano M, Pack S, Kanduri C, Kanduri M, Ginjala V, Vostrov A, Quitschke W, Chernukhin I, Klenova E, Lobanenkov V, Ohlsson R (2004) The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. Genome Res 14:1594-1602

Murrell A, Heeson S, Reik W (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat Genet 36:889-893

Nakada Y, Taniura H, Uetsuki T, Inazawa J, Yoshikawa K (1998) The human chromosomal gene for necdin, a neuronal growth suppressor, in the Prader-Willi syndrome deletion region. Gene 213:65-72

Nakamura T, Mori T, Tada S, Krajewski W, Rozovskaia T, Wassell R, Dubois G, Mazo A, Croce CM, Canaani E (2002) ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol Cell 10:1119-1128

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292:110-113

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393:386-389

Ng HH, Ciccone DN, Morshead KB, Oettinger MA, Struhl K (2003) Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. Proc Natl Acad Sci U S A 100:1820-1825

Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet 23:58-61

Nicholls RD, Knepper JL (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. Annu Rev Genomics Hum Genet 2:153-175

Nicholls RD, Saitoh S, Horsthemke B (1998) Imprinting in Prader-Willi and Angelman syndromes. Trends Genet 14:194-200

Ning Y, Roschke A, Christian SL, Lesser J, Sutcliffe JS, Ledbetter DH (1996) Identification of a novel paternally expressed transcript adjacent to snRPN in the Prader-Willi syndrome critical region. Genome Res 6:742-746

Nishioka K, Chuikov S, Sarma K, Erdjument-Bromage H, Allis CD, Tempst P, Reinberg D (2002) Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev 16:479-489

Noma K, Allis CD, Grewal SI (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science 293:1150-1155

Nowak SJ, Corces VG (2004) Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. Trends Genet 20:214-220

Ohlsson R, Renkawitz R, Lobanenkov V (2001) CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet 17:520-527

Ohta T, Khadake JR, Rodriguez-Jato S, Mione C, Knepper JL, Svoboda P, McCarrey JR, Schultz RM, Yang TP, Nicholls RD (2001) Imprint establishment in Prader-Willi/Angelman syndromes is controlled by NRF-1 and CTCF transcription factors at alternative promoters of SNURF-SNRPN. Am J Hum Genet 69:Suppl. A52

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247-257

Ozcelik T, Leff S, Robinson W, Donlon T, Lalande M, Sanjines E, Schinzel A, Francke U (1992) Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region. Nat Genet 2:265-269

Parekh BS, Maniatis T (1999) Virus infection leads to localized hyperacetylation of histones H3 and H4 at the IFN-beta promoter. Mol Cell 3:125-129

Paulsen M, Ferguson-Smith AC (2001) DNA methylation in genomic imprinting, development, and disease. J Pathol 195:97-110

Pedone PV, Pikaart MJ, Cerrato F, Vernucci M, Ungaro P, Bruni CB, Riccio A (1999) Role of histone acetylation and DNA methylation in the maintenance of the imprinted expression of the H19 and Igf2 genes. FEBS Lett 458:45-50

Pennacchio LA, Rubin EM (2001) Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet 2:100-109

Perez-Burgos L, Peters AH, Opravil S, Kauer M, Mechtler K, Jenuwein T (2004) Generation and characterization of methyl-lysine histone antibodies. Methods Enzymol 376:234-254

Perk J, Makedonski K, Lande L, Cedar H, Razin A, Shemer R (2002) The imprinting mechanism of the Prader-Willi/Angelman regional control center. EMBO J 21:5807-5814

Peterson CL, Laniel MA (2004) Histones and histone modifications. Curr Biol 14:R546-551

Pichaud F, Desplan C (2002) Pax genes and eye organogenesis. Curr Opin Genet Dev 12:430-434

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science 300:131-135

Pogo BG, Allfrey VG, Mirsky AE (1966) RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. Proc Natl Acad Sci U S A 55:805-812

Pugacheva EM, Tiwari VK, Abdullaev Z, Vostrov AA, Flanagan PT, Quitschke WW, Loukinov DI, Ohlsson R, Lobanenkov VV (2005) Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. Hum Mol Genet 14:953-965

Quan XJ, Denayer T, Yan J, Jafar-Nejad H, Philippi A, Lichtarge O, Vleminckx K, Hassan BA (2004) Evolution of neural precursor selection: functional divergence of proneural proteins. Development 131:1679-1689

Quandt K, Frech K, Karas H, Wingender E, Werner T (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res 23:4878-4884

Rank G, Prestel M, Paro R (2002) Transcription through intergenic chromosomal memory elements of the Drosophila bithorax complex correlates with an epigenetic switch. Mol Cell Biol 22:8026-8034

Razin A (1998) CpG methylation, chromatin structure and gene silencing-a three-way connection. EMBO J 17:4905-4908

Razin A, Riggs AD (1980) DNA methylation and gene function. Science 210:604-610

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406:593-599

Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, Gaszner M, Felsenfeld G (2002) Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc Natl Acad Sci U S A 99:6883-6888

Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. Science 293:1089-1093

Reik W, Walter J (2001) Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote. Nat Genet 27:255-256

Ren J, Lee S, Pagliardini S, Gerard M, Stewart CL, Greer JJ, Wevrick R (2003) Absence of Ndn, encoding the Prader-Willi syndrome-deleted gene necdin, results in congenital deficiency of central respiratory drive in neonatal mice. J Neurosci 23:1569-1573

Rice JC, Allis CD (2001) Code of silence. Nature 414:258-261

Robzyk K, Recht J, Osley MA (2000) Rad6-dependent ubiquitination of histone H2B in yeast. Science 287:501-504

Roopra A, Qazi R, Schoenike B, Daley TJ, Morrison JF (2004) Localized domains of g9a-mediated histone methylation are required for silencing of neuronal genes. Mol Cell 14:727-738

Rougeulle C, Cardoso C, Fontes M, Colleaux L, Lalande M (1998) An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. Nat Genet 19:15-16

Rougeulle C, Glatt H, Lalande M (1997) The Angelman syndrome candidate gene, UBE3A/E6-AP, is imprinted in brain. Nat Genet 17:14-15

Rougeulle C, Navarro P, Avner P (2003) Promoter-restricted H3 Lys 4 di-methylation is an epigenetic mark for monoallelic expression. Hum Mol Genet 12:3343-3348

Runte M, Huttenhofer A, Gross S, Kiefmann M, Horsthemke B, Buiting K (2001) The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. Hum Mol Genet 10:2687-2700

Saitoh N, Bell AC, Recillas-Targa F, West AG, Simpson M, Pikaart M, Felsenfeld G (2000) Structural and functional conservation at the boundaries of the chicken beta-globin domain. EMBO J 19:2315-2322

Saitoh S, Buiting K, Rogan PK, Buxton JL, Driscoll DJ, Arnemann J, Konig R, Malcolm S, Horsthemke B, Nicholls RD (1996) Minimal definition of the imprinting center and fixation of chromosome 15q11-q13 epigenotype by imprinting mutations. Proc Natl Acad Sci U S A 93:7811-7815

Saitoh S, Wada T (2000) Parent-of-origin specific histone acetylation and reactivation of a key imprinted gene locus in Prader-Willi syndrome. Am J Hum Genet 66:1958-1962

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J, Kouzarides T (2002) Active genes are tri-methylated at K4 of histone H3. Nature 419:407-411

Scarpulla RC (2002) Transcriptional activators and coactivators in the nuclear control of mitochondrial function in mammalian cells. Gene 286:81-89

Schmidt JV, Levorse JM, Tilghman SM (1999) Enhancer competition between H19 and Igf2 does not mediate their imprinting. Proc Natl Acad Sci U S A 96:9733-9738

Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T (2004) Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol 6:73-77

Schweizer J, Zynger D, Francke U (1999) In vivo nuclease hypersensitivity studies reveal multiple sites of parental origin-dependent differential chromatin conformation in the 150 kb SNRPN transcription unit. Hum Mol Genet 8:555-566

Shemer R, Birger Y, Riggs AD, Razin A (1997) Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. Proc Natl Acad Sci U S A 94:10267-10272

Shemer R, Hershko AY, Perk J, Mostoslavsky R, Tsuberi B, Cedar H, Buiting K, Razin A (2000) The imprinting box of the Prader-Willi/Angelman syndrome domain. Nat Genet 26:440-443

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119:941-953

Shibata S, Lee JT (2004) Tsix Transcription- versus RNA-Based Mechanisms in Xist Repression and Epigenetic Choice. Curr Biol 14:1747-1754

Shiio Y, Eisenman RN (2003) Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A 100:13225-13230

Simon P, Schott K, Williams RW, Schaeffel F (2004) Posttranscriptional regulation of the immediate-early gene EGR1 by light in the mouse retina. Eur J Neurosci 20:3371-3377

Sims RJ, 3rd, Nishioka K, Reinberg D (2003) Histone lysine methylation: a signature for chromatin function. Trends Genet 19:629-639

Sleutels F, Zwart R, Barlow DP (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 415:810-813

Smilinich NJ, Day CD, Fitzpatrick GV, Caldwell GM, Lossie AC, Cooper PR, Smallwood AC, Joyce JA, Schofield PN, Reik W, Nicholls RD, Weksberg R, Driscoll DJ, Maher ER, Shows TB, Higgins MJ (1999) A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. Proc Natl Acad Sci U S A 96:8064-8069

Smith KT, Coffee B, Reines D (2004a) Occupancy and synergistic activation of the FMR1 promoter by Nrf-1 and Sp1 in vivo. Hum Mol Genet 13:1611-1621

Smith NG, Brandstrom M, Ellegren H (2004b) Evidence for turnover of functional noncoding DNA in mammalian genome evolution. Genomics 84:806-813

Soejima H, Wagstaff J (2005) Imprinting centers, chromatin structure, and disease. J Cell Biochem 95:226-233

Stewart MD, Li J, Wong J (2005) Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. Mol Cell Biol 25:2525-2538

Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403:41-45

Strahl BD, Ohba R, Cook RG, Allis CD (1999) Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. Proc Natl Acad Sci U S A 96:14967-14972

Sun ZW, Allis CD (2002) Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature 418:104-108

Surani MA, Barton SC, Norris ML (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature 308:548-550

Sutcliffe JS, Han M, Christian SL, Ledbetter DH (1997) Neuronally-expressed necdin gene: an imprinted candidate gene in Prader-Willi syndrome. Lancet 350:1520-1521

Sutcliffe JS, Nakao M, Christian S, Orstavik KH, Tommerup N, Ledbetter DH, Beaudet AL (1994) Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. Nat Genet 8:52-58

Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H, Shinkai Y (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev 16:1779-1791

Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y (2004) Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell 116:51-61

Takada S, Paulsen M, Tevendale M, Tsai CE, Kelsey G, Cattanach BM, Ferguson-Smith AC (2002) Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2-H19. Hum Mol Genet 11:77-86

Tamaru H, Selker EU (2001) A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature 414:277-283

Tate PH, Bird AP (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev 3:226-231

Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247-250

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680

Thorvaldsen JL, Duran KL, Bartolomei MS (1998) Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes Dev 12:3693-3702

Tudor M, Akbarian S, Chen RZ, Jaenisch R (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proc Natl Acad Sci U S A 99:15536-15541

Turner BM (2005) Reading signals on the nucleosome with a new nomenclature for modified histones. Nat Struct Mol Biol 12:110-112

Uetsuki T, Takagi K, Sugiura H, Yoshikawa K (1996) Structure and expression of the mouse necdin gene. Identification of a postmitotic neuron-restrictive core promoter. J Biol Chem 271:918-924

Umlauf D, Goto Y, Cao R, Cerqueira F, Wagschal A, Zhang Y, Feil R (2004) Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nat Genet 36:1296-1300

van Amerongen R, Nawijn M, Franca-Koh J, Zevenhoven J, van der Gulden H, Jonkers J, Berns A (2005) Frat is dispensable for canonical Wnt signaling in mammals. Genes Dev 19:425-430

Vu TH, Li T, Hoffman AR (2004) Promoter-restricted histone code, not the differentially methylated DNA regions or antisense transcripts, marks the imprinting status of IGF2R in human and mouse. Hum Mol Genet 13:2233-2245

Walsh CP, Bestor TH (1999) Cytosine methylation and mammalian development. Genes Dev 13:26-34

Watanabe T, Yoshimura A, Mishima Y, Endo Y, Shiroishi T, Koide T, Sasaki H, Asakura H, Kominami R (2000) Differential chromatin packaging of genomic imprinted regions between expressed and non-expressed alleles. Hum Mol Genet 9:3029-3035

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520-562

Weintraub H, Groudine M (1976) Chromosomal subunits in active genes have an altered conformation. Science 193:848-856

Weksberg R, Smith AC, Squire J, Sadowski P (2003) Beckwith-Wiedemann syndrome demonstrates a role for epigenetic control of normal development. Hum Mol Genet 12 Spec No 1:R61-68

Westendorf JJ, Kahler RA, Schroeder TM (2004) Wnt signaling in osteoblasts and bone diseases. Gene 341:19-39

Wevrick R, Francke U (1996) Diagnostic test for the Prader-Willi syndrome by SNRPN expression in blood. Lancet 348:1068-1069

Wevrick R, Kerns JA, Francke U (1994) Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. Hum Mol Genet 3:1877-1882

Wirth J, Back E, Huttenhofer A, Nothwang HG, Lich C, Gross S, Menzel C, Schinzel A, Kioschis P, Tommerup N, Ropers HH, Horsthemke B, Buiting K (2001) A translocation breakpoint cluster disrupts the newly defined 3' end of the SNURF-SNRPN transcription unit on chromosome 15. Hum Mol Genet 10:201-210

Xiao B, Wilson JR, Gamblin SJ (2003) SET domains and histone methylation. Curr Opin Struct Biol 13:699-705

Xin Z, Allis CD, Wagstaff J (2001) Parent-specific complementary patterns of histone H3 lysine 9 and H3 lysine 4 methylation at the Prader-Willi syndrome imprinting center. Am J Hum Genet 69:1389-1394

Xin Z, Tachibana M, Guggiari M, Heard E, Shinkai Y, Wagstaff J (2003) Role of histone methyltransferase G9a in CpG methylation of the Prader-Willi syndrome imprinting center. J Biol Chem 278:14996-15000

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402:187-191

Yang T, Adamson TE, Resnick JL, Leff S, Wevrick R, Francke U, Jenkins NA, Copeland NG, Brannan CI (1998) A mouse model for Prader-Willi syndrome imprinting-centre mutations. Nat Genet 19:25-31

Yang Y, Hu JF, Ulaner GA, Li T, Yao X, Vu TH, Hoffman AR (2003) Epigenetic regulation of Igf2/H19 imprinting at CTCF insulator binding sites. J Cell Biochem 90:1038-1055

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335-340

Yusufzai TM, Felsenfeld G (2004) The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. Proc Natl Acad Sci U S A 101:8620-8624

Zegerman P, Canas B, Pappin D, Kouzarides T (2002) Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. J Biol Chem 277:11621-11624

Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev 13:1924-1935

Zhang Y, Reinberg D (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev 15:2343-2360

Zhao H, Dean A (2004) An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic Acids Res 32:4903-4919